PARODY AND PERCEPTION: USING COGNITIVE RESEARCH TO EXPAND FAIR USE IN COPYRIGHT

Laura R. Bradford

[pages 705–770]

Abstract: This Article draws on cognitive research to examine a conflict within copyright doctrine. Scholars typically analyze unauthorized secondary use of expressive works using an economic or a free speech analysis. The former views copyrighted works primarily as products, the latter primarily as speech. Both paradigms focus on the person doing the speaking or distributing. Copyrighted works, however, protect communication between two parties, which involves both expression and understanding of that expression. This Article argues that because copyright is a right to control certain types of information, how we process information is relevant in determining copyright’s scope. By incorporating lessons from cognitive research on memory, attention, and preference, courts can formulate rules that better balance the rights of owners with the needs of consumers and new users for open engagement with expressive works. More generally, a cognitive approach to secondary use refocuses the debate from a question of what users intend to what audiences require in choosing and consuming such works. This focus is in keeping with copyright’s goal of promoting innovation to further the public good.

RISKING THE EIGHTH AMENDMENT: ARBITRARINESS, JURIES, AND DISCRETION IN CAPITAL CASES

Janet C. Hoeffel

[pages 771–824]

Abstract: This Article argues that the stalled dialogue over the U.S. Supreme Court’s administration of capital punishment suffers from a fundamental misunderstanding of the first principles of the Eighth Amendment. Although the Court in Furman v. Georgia articulated an
Eighth Amendment substantive right against the arbitrary imposition of death sentences, the Court later recast *Furman* to require procedures that merely reduced a substantial risk of arbitrariness. Instead, *Furman* mandates procedures that expose arbitrariness. The best vehicle for this is a review of jurors’ reasons for imposing death in an individual case. Although there are political and practical hurdles to mining the jurors’ reasons for imposing death, they are far from insurmountable. Absent a moratorium, this Article advocates change that informs and exposes the process of death.

**NOTES**

**STRUCTURE VERSUS EFFECT: REVEALING THE UNCONSTITUTIONAL OPERATION OF TITLE IX’S ATHLETICS PROVISIONS**

*Jennifer R. Capasso*

[pages 825–862]

**Abstract:** Title IX of the 1972 Education Amendments prohibits discrimination on the basis of sex in any educational program receiving federal financial assistance, including athletics. Although the statute has brought great improvements in female athletics, such achievements have not come without problems. Title IX’s complex regulatory framework, developed by the Department of Education, has led to numerous lawsuits. In early cases, courts interpreted and upheld the framework as consistent with Title IX. This Note argues, however, that the courts’ early decisions and interpretation of the framework, coupled with educational institutions’ current fiscal constraints, have left institutions with only one option for compliance: cutting men’s teams. Such an inflexible framework amounts to a gender-based quota system that, under recent affirmative action caselaw, courts should hold unconstitutional. Only then will the Department of Education have the impetus to reinterpret the regulatory framework so that Title IX can operate in the way in which it was originally intended.

**INTEREST, PRINCIPAL, AND CONCEPTUAL SEVERANCE**

*Rebecca Rogers*

[pages 863–891]
Abstract: When considering applications of the Takings Clause to appropriations of interest and principal, courts rely on the quaintly expressed 18th century rule that “interest shall follow the principal, as the shadow the body.” Since that time, an elaborate body of regulatory takings law has evolved. As part of this evolution, courts delved into questions of temporary takings of property, occurring both as a result of outright physical occupation and through regulatory limitations. This Note argues that courts should reject the old “interest follows principal” rule and adopt the modern temporary takings rules. In particular, this Note explores the results of such a change in three cases involving prisoners alleging an unconstitutional taking when their interest was deposited in accounts with the Departments of Corrections while incarcerated. Adopting a temporary takings analysis would not only bring consistency between the analyses of takings of money and takings of land, but also offer a more consistent outcome within this group of prisoner interest cases.

PUBLIC EMPLOYEE FREE SPEECH: THE POLICY REASONS FOR REJECTING A PER SE RULE PRECLUDING SPEECH RIGHTS

Marni M. Zack

[pages 893–919]

Abstract: Public employees do not enjoy the same free speech rights under the First Amendment as do ordinary citizens. The Fourth Circuit Court of Appeals recently adopted a virtual per se rule precluding free speech rights for public employees while they are performing ordinary job duties. This Note argues that such a per se rule both lacks policy justification and, more importantly, would undermine the purposes of the First Amendment by impeding academic freedom and permitting viewpoint discrimination. Rejecting the per se rule best preserves the free speech rights of public employees and in turn allows them to speak freely about governmental operations—an especially important function of public employees, who are firsthand witnesses to government activities.
PARODY AND PERCEPTION: USING COGNITIVE RESEARCH TO EXPAND FAIR USE IN COPYRIGHT

Laura R. Bradford*

Abstract: This Article draws on cognitive research to examine a conflict within copyright doctrine. Scholars typically analyze unauthorized secondary use of expressive works using an economic or a free speech analysis. The former views copyrighted works primarily as products, the latter primarily as speech. Both paradigms focus on the person doing the speaking or distributing. Copyrighted works, however, protect communication between two parties, which involves both expression and understanding of that expression. This Article argues that because copyright is a right to control certain types of information, how we process information is relevant in determining copyright’s scope. By incorporating lessons from cognitive research on memory, attention, and preference, courts can formulate rules that better balance the rights of owners with the needs of consumers and new users for open engagement with expressive works. More generally, a cognitive approach to secondary use refocuses the debate from a question of what users intend to what audiences require in choosing and consuming such works. This focus is in keeping with copyright’s goal of promoting innovation to further the public good.

Introduction

In the fall of 2004, Little, Brown and Company (“Little Brown”) published Yiddish with Dick and Jane. As in the original, Dick and Jane teach basic language skills, except that in this version they are adults, the words are Yiddish, and they live in a suburban neighborhood rife with adultery and drug use. Pearson, the publisher of the original

* © 2005, Laura R. Bradford, Frank H. Marks Visiting Associate Professor and Administrative Fellow in the Intellectual Property Program, George Washington University Law School. I would like to thank David Kirkpatrick, Niva Elkin-Koren, Diane Zimmerman, Michael Abramowicz, Rebecca Hollander-Blumoff, Kerry Abrams, Mark A. Lemley, Laura Heymann, Mark McKenna, the Lawyering Colloquium, the NYU Law and Humanities Colloquium, and the MSU Young IP Scholars Workshop for their thoughts and comments. Alexis Stone and Zach Winnick provided valuable research assistance. Any remaining errors are my own. The author grants others permission to make copies of this article for use by nonprofit educational institutions, so long as the copies are distributed at or below cost, the author and journal are identified, and proper notice of copyright is affixed to each copy.
Dick and Jane books, has not expressed interest in licensing a Yiddish version. But in January 2005 Pearson sued Little Brown for, among other things, copyright infringement because the new portrayal of Dick and Jane is unsuited for association with the first grade reading primer and the nostalgic memories it inspires.\(^1\) Given that a \textit{Fun with Dick and Jane} movie is scheduled for release in December 2005, Pearson would rather not alienate its potential audience.\(^2\) Little Brown, publisher of the Yiddish version, claims the book is social commentary entitled to First Amendment protection.\(^3\)

Pearson’s real concern is for the perception of its characters among consumers and this Article closely examines such claims. Some might argue that concern about consumer reactions does not belong within copyright law. Trademark law, with its prohibitions against blurring and tarnishment of symbols, might seem better suited to these issues and many plaintiffs, including Pearson, file suit under both theories. Copyright law, however, also looks to consumer perceptions to set boundaries around protected works and determine what unauthorized uses qualify as “fair.”

Reliance on consumer reactions is pervasive in copyright law. Modern copyright law forbids use not only of exact text, but also of any communication audiences recognize as “substantially similar” or obviously derived from an original. Infringement exists whenever an ordinary observer would conclude that the defendant has incorporated something of substance and value from the plaintiff’s work, regardless of whether audiences would mistake the second work for the former.\(^4\) New uses are “fair” only to the extent they do not adversely

\(^2\) Id., ¶ 30; see also Edward Wyatt, Primer Spoof with Yiddish Faces Suit (in English), N.Y. Times, Jan. 15, 2005, at B7.
\(^3\) Wyatt, supra note 2, at B7.
\(^4\) See, e.g., Incredible Techs., Inc., v. Virtual Techs., Inc., 400 F.3d 1007, 1011 (7th Cir. 2005) (describing the “ordinary observer” test in the context of video games as “whether the accused work is so similar to the plaintiff’s work that an ordinary reasonable person would conclude that the defendant unlawfully appropriated the plaintiff’s protectible expression by taking material of substance and value.’’); infra notes 179–83 and accompanying text. It is worth noting that general exceptions such as the merger doctrine and scenes a faire exist. The scenes a faire doctrine exempts stock scenes and incidents that are common to certain genres or themes, such as car chases in police movies or depictions of killing in vampire stories, from being claimed by any one party. 4 Melville B. Nimmer & David Nimmer, Nimmer on Copyright § 13.03[B][4] at 13-78.6 to -78.7 (perm. ed., rev. vol. 2005). The merger doctrine states that when an idea and its expression have become synonymous, so that there is no other way to express the idea, the expression itself is unprotected. \textit{Id.} § 13.03[B][3] (listing as examples a label depicting a piece of fruit or a map
affect the market or “value” of the original—a test that asks courts to consider consumer appeal and demand. Such amorphous tests already grant owners control over most public representations of their works, regardless of whether one thinks copyright law should extend to issues of consumer perception. Accordingly, as court decisions have narrowed trademark dilution protection, producers can increasingly rely on copyright’s broader provisions to protect songs, characters, images, and texts against dilution of meaning and value in the face of imaginative reworkings of popular fare, or “secondary use.”

Pearson’s complaint echoes the typical claim for such broad copyright protection: that if courts allow too much secondary use of expressive works without the owner’s permission, the original works themselves will be degraded and overexploited. Owners of expressive works claim loss of control over the presentation of a work, be it an image, film, character, or song, has the potential to destroy the public’s positive associations with the original and so exhaust the demand for the original and its attendant products. Such use also allows new users to “free-ride” on the owner’s investments in building an audience for the original work.

This harm is typically analogized to the problem of “overgrazing” in natural resource law. This paradigm suggests that those with

---

5 17 U.S.C. § 107(4) (2000); see also infra notes 251–52 and accompanying text.
8 See Lawrence Lessig, Re-Marking the Progress in Frischmann, 89 MINN. L. REV. 1031, 1034 (2005) (describing copyright as preventing free-riding by competitors); Richard A. Posner, When is Parody Fair Use?, 21 J. LEGAL STUD. 67, 72 (1992) (arguing that allowing free use of older works redirects resources to copying and reuse, leading to under-investment in production).
9 See Lee Anne Fennell, Common Interest Tragedies, 98 NW. U. L. REV. 907, 918–19 (2004) (stating that overgrazing can occur even with nonrivalrous goods such as songs and theo-
shared access to a resource have no incentive to maintain its value and so will overuse it until depletion. Only a private right to exclude others will solve this “tragedy of the commons.”

Academic critiques of using an overgrazing doctrine for intellectual property are widespread. Critics claim that copyright’s purpose of encouraging investment in creative works does not require control over all manifestations of that work after creation. It is unclear to what extent multiple versions harm or degrade the audience’s ability to connect with the original, thereby altering incentives to produce creative works. Increasingly, First Amendment scholars and audience members resent that any one person should own the right to communicate alternative conceptions of an expressive work.

These critiques recognize that the problem with importing an overgrazing doctrine into copyright is that expressive property does not behave exactly like real property. Where real property gives owners rights to a demarcated piece of ground, the precise boundaries of a right to tell a story or perform a song are difficult to measure. Unlike rights in tangible objects, copyright grants rights in something more ephemeral: information and communication.

Digital technology exacerbates the tension between audience and owner in two ways. First, it facilitates copying, alteration, and distribution of copyrighted works. Second, it makes tangible informal communication and social interaction through email, websites and other forms of electronic communication, and so pulls relatively private speech into the reach of the copyright laws.

While rejection of overgrazing might be fashionable in academia, courts have endeavored to find a middle ground. Through use of a “parody” test, courts seek to allow important critical speech to reach the marketplace, but prevent free-riding or use of old works to get atten-

---


14 For further discussion, see text accompanying notes 200–02. See also Lessig, supra note 13, at 8.
tion for unrelated works.\textsuperscript{15} This test allows speech directly commenting on previous work to use the previous work free of charge. If the second work uses the older work for attention, or to satirize society in general, the second user must get permission or risk injunction.\textsuperscript{16} The difficulty, however, of locating a precise point of view in ambiguous creative works creates an uncertain standard that chills potential speakers and discourages investment in satirical works. Small-scale users are especially disadvantaged because they generally lack the resources to bargain for licenses or defend potential lawsuits.\textsuperscript{17}

This Article argues that by paying attention to the significant body of research about how people process and use information, courts can formulate more tailored rules regarding use and reuse of informational products, specifically cultural works such as books, movies, characters, and songs. Because copyright is a right to control the flow of certain types of information, how we process information is relevant in determining copyright’s scope. Other areas of intellectual property law, such as trademark, implicitly recognize that rights in information depend on not only the content of the information but also its effect on the recipient. Only uses that cause consumer confusion or demonstrably alter consumer perception are actionable in trademark. Recognition of similar substantive limitations in copyright law has lagged.

By acknowledging the dynamic properties of information as opposed to viewing it as a tangible entitlement, we can locate a more exact balance between the rights of owners and the needs of users and consumers for diverse and open engagement with expressive works. Cognitive research on memory, attention, and preference suggests that, as a practical matter, we should provide more leeway to secondary uses of expressive works that exhibit characteristics such as clear source information and audience opt-in distribution methods. As a normative matter, cognitive research indicates that we should be concerned about the consumer’s ability to choose between competing interpretations of a work, not the owner’s authority to control public


\textsuperscript{16} See infra notes 66–68, 257–60, and accompanying text.

\textsuperscript{17} See Molly Shaffer Van Houweling, Distributive Values in Copyright, 83 Tex. L. Rev. 1536, 1564–67 (2005). Such users are also disadvantaged by the considerable administrative burden of locating copyright owners for older works. This burden has grown steeper in recent years because copyright’s extended term means that original authors may be long-dead and heirs and transferees scattered or forgotten. In addition, notice, registration, and periodic renewal, which ensured information about copyright owners remained publicly available, are no longer required for copyright protection.
perception. This approach conserves the owner’s incentive to invest in authorized reproductions while allowing greater participation by outsiders in reforming, debating, and refining the work.\textsuperscript{18}

In other areas of the law, research from cognitive and behavioral psychology has provided insights on how to distribute resources efficiently and to resolve private disputes. Studies of human perception and decision-making faculties have been used to reframe debates about tort remedies and contract formation.\textsuperscript{19} In property law, acknowledgement of decision-making biases has prompted re-examination of the idea that broad, well-defined property rights facilitate rational and efficient market exchanges.\textsuperscript{20} As the nature of property grows more intangible and the nature of ownership more diffuse, scholars are looking at how common cognitive biases and processes impact the optimal allocation of entitlement rights in a variety of settings.\textsuperscript{21} To date, however, the cognitive components of certain forms of intellectual property such as copyright, trademark, or rights of publicity have received little to no academic attention. This void is curious given that the boundaries of the entitlements themselves reference human perception.\textsuperscript{22}

With respect to copyright in particular, attention to cognitive psychology and consumer research can provide more tailored rules for secondary use. Copyright’s purpose is to enrich the public by encourag-
ing the production of creative works. The public is best served by encouraging greater numbers of innovative uses until such uses distort consumers’ ability to identify and choose between works. Such an approach removes the emphasis on the motives of the secondary user, as under the current fair use test, and focuses instead on the needs of the consumer. Well-established doctrines from social cognition research indicate that we can better distinguish between the types of secondary uses that undermine consumer choice and those that do not.

Four doctrines from cognitive science—attitude resistance, source effects, frequency effects, and hierarchy of processing research—provide an illustration. Taken together they suggest that certain uses of copyrighted materials are unlikely to harm materially an owner’s investment in, or a consumer’s ability to, distinguish a specific message or identity in a given informational work. Recognition of such information processing effects would allow more leeway for secondary use that exhibits characteristics such as clear source information and audience opt-in distribution methods. Such doctrines also suggest more freedom for secondary use of older rather than newer works. To offer two simple examples, a literary parody of Dick and Jane, clearly marked as such and available only to willing purchasers, does not seem likely to do much harm to the general public’s ability to recognize authorized versions of the characters. By contrast, a ubiquitous television commercial using Dick and Jane to promote Viagra may unconsciously distort audience memory and attitude. Such distortions unduly interfere with the owner’s ability to build an audience for a communicative good. Under the current parody regime, determination of whether a given use is fair can only be made after distribution and the filing of a lawsuit. The guidelines outlined above could facilitate the creation of more definitive safe harbors for certain derivative uses within fair use and so obviate the need for a separate fair use category for “parody.”

Part I of this Article looks at economic and free speech accounts of secondary use and shows how each omits consideration of elements that are important to consumers. Part II explains the reasons why a

23 See Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 432 (1984) (stating that the ultimate aim and sole interest of the Copyright Act is to stimulate artistic creativity for the public good); accord Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975). But see Eldred v. Ashcroft, 537 U.S. 186, 212 n.18 (2003) (“[C]opyright law serves public ends by providing individuals with an incentive to pursue private ones.”).

24 See Aiken, 422 U.S. at 156 (“Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts.”).

25 See infra notes 31–131 and accompanying text.
cognitive analysis is relevant to copyright law.\textsuperscript{26} It traces the historical development of copyright from a right to reproduce tangible objects to a right of conceptual representation and shows how contemporary debates over branding and reuse stem from this expansion.\textsuperscript{27} Part III introduces a new standard for secondary use based on cognitive principles that could replace, or at least augment, the current problematic parody analysis.\textsuperscript{28}

First, however, an important word about the vocabulary within this Article. Cognitive psychology refers to the study of how people process and remember information.\textsuperscript{29} Related areas in social cognitive research examine how the processing of information affects attitudes and preferences.\textsuperscript{30} At the risk of muddying the social science waters, I use the term “cognitive research” to refer to both areas of scientific study. I also draw on behavioral and ethnographic research about marketing and consumer preference, using the umbrella label “consumer research” to refer to this body of knowledge.

I. PREVAILING ACCOUNTS OF SECONDARY USE: TRESPASS AND FREE SPEECH

A. The Economic Account of Expressive Property

The prevailing account of copyright and secondary use treats expressive property like real property. This framework, drawn from Law and Economics scholarship, emphasizes the importance of private ownership as the solution to “the tragedy of the commons” problem. In this view, common ownership of resources is inefficient because each individual stakeholder has little incentive to maintain or improve the resource but instead will overuse it. Private ownership, on the other hand, provides incentives to improve and maintain the property and allows for full internalization of the costs of different choices. Applying this analysis to intellectual property frames all unauthorized

\textsuperscript{26} See infra notes 132–321 and accompanying text.
\textsuperscript{27} See infra notes 132–270 and accompanying text.
\textsuperscript{28} See infra notes 322–71 and accompanying text.
secondary users as trespassers and their secondary works as stealing.\textsuperscript{31} The problem with this analysis is that real property insights based on the behavior of farmers in pastures are of limited use in an intangible format grounded in human perception.\textsuperscript{32}

Under the tragedy of the commons model, in the absence of private property rights, users will overexploit a pasture because, without a guarantee that the pasture will be properly managed over time, each individual user’s best interest is to “overgraze” or take as much grass as possible for their herd.\textsuperscript{33} The same insight has been made about hunting: where no one controls the ability of others to hunt, no one has incentive to increase or maintain the stock of game.\textsuperscript{34} In either case, the richness of the soil or the stock of game will quickly diminish.

Private ownership, by contrast, allows for consideration of the costs and benefits of different uses of a resource.\textsuperscript{35} Owners will allocate rights to the land to maximize its value and increase their own wealth or happiness.\textsuperscript{36} Private ownership will induce investment in improving the property because owners know they will be able to capture the return on their investment.\textsuperscript{37}

Concentrated ownership of land also lowers the transaction costs of private agreements. Where ownership is concentrated, outside parties may easily contract for any beneficial use of the property.\textsuperscript{38} The identity of the initial owner matters little then; in an efficient market with low transaction costs, the owner will sell or license rights to the party who will put the parcel to its highest and best use.\textsuperscript{39} The economic approach presumes that consensual market transfers ordinarily

\begin{footnotes}
\footnotetext{31}{See Posner, \textit{supra} note 8, at 67, 72.}
\footnotetext{32}{Others have criticized the tendency to draw direct analogies between real and intangible property. \textit{See}, \textit{e.g.}, Lemley, \textit{Free-Riding, supra} note 11, at 2–3; William F. Patry \& Richard A. Posner, \textit{Fair Use and Statutory Reform in the Wake of Eldred}, 92 \textit{Cal. L. Rev.} 1639, 1643 (2004). These critiques focus on the differences between real and intangible property from an economic perspective. I also am interested in these differences, but in this Article I explore additional perspectives beyond economics to describe the properties of intangible goods.}
\footnotetext{33}{Harold Demsetz, \textit{Toward a Theory of Property Rights}, 57 \textit{Am. Econ. Rev.} 347, 354 (1967); Landes \& Posner, \textit{supra} note 7, at 484.}
\footnotetext{34}{Demsetz, \textit{supra} note 33, at 351.}
\footnotetext{35}{\textit{See id.} at 348.}
\footnotetext{36}{\textit{Id.} at 355.}
\footnotetext{37}{\textit{See id.} at 356.}
\footnotetext{38}{\textit{Id.} at 357.}
\footnotetext{39}{Ronald Coase, \textit{The Problem of Social Cost}, 3 \textit{J.L. \& Econ.} 1, 7, 11 (1960).}
\end{footnotes}
are the most effective way to distribute resources; departures from this norm in the form of legal regulation bear a heavy burden of proof.\(^{40}\)

Scholars apply the tragedy of the commons model broadly with the aim of creating rules that stretch across different types of entitlements. They do so on the theory that a property doctrine that applies universally promotes certainty in contracting and lowers information costs for businesses.\(^{41}\) Legal exceptions to broad property regimes frustrate competition and promote rent-seeking behavior among interest groups.\(^{42}\)

According to this unified theory, incentives operate similarly with intellectual entitlements, such as copyrights, trademarks, and the right of publicity, as they do with tangible property.\(^{43}\) Private ownership creates rewards for producing expressive property.\(^{44}\) The right to exclude others from free-riding on this effort creates incentives to maintain and invest in creation and spurs further creative efforts.\(^{45}\) Owners of intangible rights will choose licensing and distribution arrangements that will maximize the value of the right.\(^{46}\) In a costless


\(^{42}\) Easterbrook, supra note 41, at 107–08.

\(^{43}\) See, e.g., Frank H. Easterbrook, Intellectual Property is Still Property, 13 HARV. J.L. & PUB. POL’Y 108, 112 (1990) ("[A] right to exclude in intellectual property is no different in principle from the right to exclude in physical property."); Landes & Posner, supra note 7, at 484; see also Eldred v. Ashcroft, 537 U.S. 186, 212 n.18 (2003) (stating that the economic philosophy behind the Copyright Clause is to provide material incentives to individuals to produce works of knowledge).

\(^{44}\) Eldred, 537 U.S. at 212 n.18.


\(^{46}\) See, e.g., William M. Landes, Copyright, Borrowed Images, and Appropriation Art: An Economic Approach, 9 GEO. MASON L. REV. 1, 6 (2000); Merges, Are You Making Fun of Me?, supra note 40, at 306 (“[C]onsensual market transfers are the most effective way to simultaneously pursue the twin goals of incentive and dissemination.”); Posner, Parody, supra note 8, at 69 (stating that even where a potential derivative use of copyright is better or more valuable, transaction costs will be minimized by concentrating ownership in a single pair of hands).
environment, owners will also choose the uses most beneficial to society because it will be economically efficient to do so.\textsuperscript{47}

But what of overgrazing? Expressive property is intangible and so may be less vulnerable to overuse than soil or hunting grounds. Further, its consumption is “nonrivalrous.”\textsuperscript{48} Use of a book, movie, image, or other form of commodified information does not impose a direct cost on others using the same information.\textsuperscript{49} Expressive works may be endlessly copied without diminishing the quality of the original.\textsuperscript{50}

Economic theorists maintain that intellectual property may nonetheless be vulnerable to overgrazing. For example, trademarks are vulnerable to dilution when used even by non-competitors to describe unrelated goods.\textsuperscript{51} The use of “Gucci” on a diaper bag may lessen the value of the mark by weakening its connection to luxury goods.\textsuperscript{52} Similarly, commercial use of a celebrity’s image in connection with a variety of different products is likely to lessen the value of that celebrity’s endorsement.\textsuperscript{53}

Producers claim that the same effect appears in copyright. Mattel polices unauthorized uses of its Barbie character to avoid its use in art works or pop songs that might displease or disgust its target audience of young girls and their mothers.\textsuperscript{54} As William Landes and Richard Posner have noted, Disney actively manages its characters to avoid overkill and audience confusion.\textsuperscript{55} They observe:

\textsuperscript{47} Posner, supra note 8, at 73 (stating that in all but the case of direct criticism, the market is best placed to make an efficient trade-off between the First Amendment value of a parody and the cost to the social value of the original work); see also Paul Goldstein, Copyright’s Highway 178–79 (1994).

\textsuperscript{48} Lemley, Ex Ante Versus Ex Post, supra note 11, at 143.

\textsuperscript{49} Lemley, Free-Riding, supra note 11, at 25.


\textsuperscript{51} Landes & Posner, supra note 7, at 485.


\textsuperscript{53} Landes & Posner, supra note 45, at 362–63; Stuart Elliott, Critics Claim Multiple Deals Risk Saturation, USA TODAY, Apr. 30, 1991, at 1B.

\textsuperscript{54} Mattel, Inc. v. Walking Mountain Prods., 353 F.3d 792, 796, 802 (9th Cir. 2003); Mattel, Inc. v. MCA Records, Inc., 296 F.3d 894, 899 (9th Cir. 2002) (suing in trademark); Mattel, Inc. v. Pitt, 229 F. Supp. 2d 315, 318–19 (S.D.N.Y. 2002).

\textsuperscript{55} Landes & Posner, supra note 7, at 486 (“To avoid overkill, Disney manages its character portfolio with care. It has hundreds of characters on its books, many of them just waiting to be called out of retirement . . . . Disney practices good husbandry of its characters and extends the life of its brands by not over-exposing them . . . . They avoid debasing the currency.”) (alteration in original).
If . . . anyone were free to incorporate the Mickey Mouse character in a book, movie, song etc., the value of the character might plummet. Not only would the public rapidly tire of Mickey Mouse, but his image would be blurred, as some authors portrayed him as a Casanova, others as catmeat, others as an animal-rights advocate, still others as the hen-pecked husband of Minnie.  

Without strong ownership rights, those seeking to exploit the value created and maintained by a work’s owners will exhaust the work’s potential without considering the cost of their actions to others or to the work itself. Thus, economic theorists contend that the need of owners of expressive property to exclude trespassers is as great for expressive works as it is for owners of pastureland or hunting grounds.

According to the economic account, the need to exclude may be even greater with respect to expressive works because of the particular dynamics of the media and entertainment industries. Many works of authorship take years to create. The author thus incurs substantial costs, at least in the form of opportunity costs. Publishers too must decide whether to distribute a book before knowing whether it will succeed. They commit huge sums in advance to distribution and marketing, but only a very few expressive works earn back these initial investments. These blockbusters subsidize the creation and marketing of all of the other works that never find an audience. Producers thus need to recoup all of the social value of successful works to continue to produce new works. Any free-riding on their investment will redirect resources to copycats and secondary use, where creators can make use of already successful works for free instead of engaging in the risky business of creating works from scratch.

These concerns underlie the claim that the list of exceptions allowed for secondary use of expressive property should be narrow. Economic scholars argue that fair use exemptions are justified mainly in limited circumstances of market failure. For example, because

---

56 Landes & Posner, supra note 7, at 487–88; see Hughes, supra note 7, at 926, 941–42.
57 Landes & Posner, supra note 7, at 484.
58 Id. at 485.
59 Id. at 495.
60 Posner, supra note 8, at 72.
61 Id.; Patry & Posner, supra note 32, at 1644.
62 See, e.g., Goldstein, supra note 47, at 224, 230 (arguing that as technology reduces transaction costs, the need for safety valves like fair use declines, but excepting limited distribitional goals such as education and research); Tom W. Bell, Fair Use vs. Fared Use:
few owners will willingly submit their works to ridicule, private transactions will fail to create a licensing market for parody.\textsuperscript{63} Even if the social value of the parody is great, the transaction costs, here in the form of the owner’s reputational interests, are too high to permit efficient functioning of the market.\textsuperscript{64} But to prevent rampant free-riding under the guise of parody, economic theorists contend that the parody exception should only apply where the original work is unquestionably the target of the second author’s criticism.\textsuperscript{65}

By contrast, secondary users wishing to make use of a text as a vehicle or “weapon” to criticize society generally should be able to bargain for a license in the competitive market.\textsuperscript{66} So long as such users pay a market price, the theory goes, owners will be happy to facilitate this wider dissemination of their products.\textsuperscript{67} Therefore, users need no fair use exception to remove market obstructions.\textsuperscript{68}

The Supreme Court has embraced this framework in two landmark decisions governing secondary use. First, in 1985, in \textit{Harper \& Row Publishers, Inc. v. Nation Enterprises}, the Court looked at whether “the user stands to profit from exploitation of the copyrighted material without paying the customary price”—in other words, whether a finding of fair use would preclude efficient market transactions—in finding that publication of a small excerpt from a newly released presidential memoir was not fair use.\textsuperscript{69} Later, in 1994, in \textit{Campbell v. Acuff-Rose Music, Inc.}, which concerned a rap version of a famous fifties bal-

\begin{thebibliography}{99}
\bibitem{63} See Posner, \textit{supra} note 8, at 73–74.
\bibitem{64} See id.
\bibitem{65} Id. at 71; Paul Goldstein, \textit{Derivative Rights and Derivative Works in Copyright}, 30 J. Copyright Soc’y 209, 235 (1983).
\bibitem{66} Posner, \textit{supra} note 8, at 71 (describing the purpose of many parodies as simple amusement); see Goldstein, \textit{supra} note 65, at 235.
\bibitem{67} See Posner, \textit{supra} note 8, at 72–73; Goldstein, \textit{supra} note 65, at 216.
\bibitem{68} See Posner, \textit{supra} note 8, at 73.
\bibitem{69} 471 U.S. 539, 562 (1985).
\end{thebibliography}
lad *Pretty Woman*, the Court limited the parody exception to cases of
direct commentary on a theory of market failure.\(^70\) The Court distin-
guished between works that use “elements of a prior author’s compo-
sition to create a new one that, at least in part, comments on that au-
thor’s works” with those where the “commentary has no critical bearing
on the . . . original composition, which the alleged infringer merely
uses to get attention or to avoid the drudgery in working up something
fresh . . . .”\(^71\) The Court reasoned that because creators of imaginative
works would never license critical lampoons of their own productions,
judicial intervention was necessary to further valuable social commen-
tary.\(^72\) Where the second work used the original to criticize something
else, the Court ruled it could “stand on its own two feet and so re-
quire[d] justification for the very act of borrowing.”\(^73\)

This economic view of intellectual property is evident in more
recent decisions. In 2004, faced with the question of whether a sam-
pling of musical notes constituted infringement even if no one could
recognize the similarity, the Sixth Circuit Court of Appeals in *Bridge-
port Music v. Dimension Films* considered which result would lead to the
most efficient stewardship of musical recordings and ruled the sam-
pling to be infringement.\(^74\) Continuing the analogy of copyright with
real property, the court stated even the borrowing of a small number
of musical notes should be analyzed as “a physical taking rather than
an intellectual one.”\(^75\)

The *Bridgeport* decision incorporates the common mistake that all
secondary use of copyrighted works is analogous to a physical trespass

\(^71\) Id. at 580.
\(^72\) See id. at 580–81.
\(^73\) Id. at 581.
\(^75\) Id. The court’s decision turned on the different treatment of “sound recordings”
and “musical compositions” in the Copyright Act. See id.; see also 17 U.S.C. § 114(b) (2000)
(stating that the owner of a copyrighted sound recording has the exclusive right “to pre-
pare a derivative work in which the actual sounds fixed in the sound recording are rear-
anged, remixed, or otherwise altered in sequence or quality”). Because the Copyright Act
grants owners the exclusive right to “alter” sound recordings, the court concluded that the
substantial similarity test did not apply. Bridgeport Music, 383 F.3d at 398 n.8. It is not clear,
however, that the Copyright Act grants owners rights over alterations beyond recognition.
The court rejected the idea of a de minimus exception because of a disdain for free-riders;
it held that “even when a small part of a sound recording is sampled, the part taken is
something of value,” even if the only value was to “save costs” for the secondary user. Id. at
399. Such an analysis ignores the distributional concerns underlying the fair use exception
without explaining how a de minimus exception would materially impact copyright’s in-
centive functions.
in real property. Bright-line rules for real property create market efficiency but differences between real estate and creative works alter where those lines should be drawn with respect to creative works. Until we understand what we mean to protect with copyright, we cannot understand what constitutes a trespass of that right.

Insights about real property fail to allocate rights efficiently in the case of intangible works in part because of differences in the nature of the entitlement. The ability to unbundle rights to real estate is generally well understood. Acknowledgment of the dynamics of real property pervades the default rules around which parties license, such as the liability rules for removal of subjacent support for example.76 With expressive property, these dynamics remain mysterious and under dispute. On the one hand, consumption of expressive goods is nonrivalrous. One’s use of one copy of a text will not affect another’s ability to consume another copy. On the other hand, the overgrazing paradigm suggests the opposite: that any secondary use of a character, image, tune, or design affects the perception of the work as a whole and so its market value. Individual members of ownership entities tend to be risk-averse when confronting this uncertainty. The prevailing wisdom in the entertainment industry favors uniformity of message across uses of protected characters and narratives to prevent expressive “overgrazing.”77 Thus owners have a powerful disincentive to license use of their properties to all but a narrow segment of goods the owner identifies as complementary.78 This set of products is likely to be much less than the set of potential products that audience members might find valuable, useful, or interesting.

The scalability of information products, which enables them to serve both massive and very small numbers of users, also prevents beneficial private transactions. In a real property circumstance, the tangible nature of the goods establishes a certain scale of potential uses. Parties seeking licenses to and easements on a three-acre parcel tend to be limited in number, easily identifiable, and represent a certain subset of activities consistent with the character and size of the land.79 Expres-

---

76 See Platts v. Sacramento N. Ry., 253 Cal. Rptr. 269, 271–72 (1988) (owner of subjacent support is absolutely liable for damages caused to surface owner by removal of the natural necessary support; liability survives even after remover has sold the property).
77 See infra notes 207–41 and accompanying discussion.
78 See Guy Pessach, Copyright Law as a Silencing Restriction on Noninfringing Materials: Unveiling the Scope of Copyright’s Diversity Externalities, 76 S. Cal. L. Rev. 1067, 1089–90 (2003) (arguing that economies of scale lead corporate media to produce homogenous content and neglect minority tastes).
79 See Patry & Posner, supra note 32, at 1643.
sive property by contrast is endlessly multipliable, may be distributed in numerous formats, and appeals to both large and small-scale users simultaneously. The concentration of ownership in a single party reduces transaction costs for large-scale licensees, but imposes additional burdens on smaller users. Individual authors and small entities often lack the information and resources necessary to engage in negotiations over established properties. They do not belong to the proper networks, and may not be sufficiently repeat players to invest in joining them. For example, many large film studios will not accept story or product pitches unless they come from an established agency or other known industry channels. In addition, the potential revenue from such licenses is so small that it is usually not worth the owner’s time to engage in negotiations. Disney, for one, will not consider a new product line unless it has the potential to become a billion-dollar revenue stream. In this way, the economic view tends to concentrate control over the dissemination and social construction of expressive works in the hands of a few large and conformist organizations.

Moreover, the ambiguity surrounding what exactly constitutes a parody creates uncertain boundaries for innovation using older texts. As discussed in Part II, the impulse to discuss and reinterpret popular texts is ancient and widespread. Examples may be found among the plays of the ancient Greeks, in parodies of early Jewish texts, and in Chaucer’s *Canterbury Tales*. The Supreme Court’s definition of parody as something that targets a specific work is somewhat narrower than the historical definition, which identifies parody as any work that adopts

---

81 See Al Lieberman & Patricia Esgate, The Entertainment Marketing Revolution: Bringing the Moguls, the Media, and the Magic to the World 5 (2002); infra notes 132–321.
82 Parody is a form of satire, in which prevailing vices or follies are held up to ridicule, but is particular in that parody incorporates material from a target text as a constituent part. Margaret A. Rose, Parody: Ancient, Modern & Post-Modern 80–81 (1993).
84 Rose, supra note 82, at 120.
the style or substance of an original for comic effect or ridicule.\textsuperscript{86} Although parody is a derivative art form, some scholars ascribe to it a significant role in breaking down old forms and ushering in new styles. At least one theorist gives it a dominant role in the creation of the novel in Europe because parody accustomed audiences to works of fiction in which the author’s voice was concealed behind the voices of others.\textsuperscript{87} Parody has also historically been a favorite device of young authors as they try out different styles before developing their own.\textsuperscript{88}

The narrow exception for targeted criticism endangers this tradition. Judges applying the parody test struggle to isolate one prevailing “message” from a complex expressive work.\textsuperscript{89} No case explains a certain process by which a judge should reach a conclusion about a work’s message, and whether that message “comments” on a previous work, other than presumably by instinct and intuition.\textsuperscript{90} As Rebecca Tushnet has pointed out, courts may be incapable of making such a determination.\textsuperscript{91} To borrow her example, whether Warhol’s paintings of Campbell’s soup cans “commented” on the cans or used them just for attention is a question capable of intense debate with no satisfactory answer.\textsuperscript{92} The resulting uncertainty about the legal standard coupled with the property remedy of injunction for violation chills potential speakers and discourages investment in satirical works.\textsuperscript{93}

Furthermore, by the time judges, who disproportionately represent a few limited segments of society, can perceive a parodic message, the social flaw being exposed is generally well-understood. Judges have tended to be more lenient to works that speak across well-known

\textsuperscript{86} Rose, \textit{supra} note 82, at 81–82.

\textsuperscript{87} Mikhail Mikhailovich Bakhtin, \textit{The Dialogic Imagination: Four Essays} 6, 309 (Michael Holquist ed., Caryl Emerson & Michael Holquist trans., Univ. of Tex. Press 1981); \textit{see also} Rose, \textit{supra} note 82, at 132.

\textsuperscript{88} Rose, \textit{supra} note 82, at 30; \textit{see} Harold Bloom, \textit{The Anxiety of Influence: A Theory of Poetry} 30 (1975).

\textsuperscript{89} \textit{See} Rebecca Tushnet, Copyright as a Model for Free Speech Law: What Copyright Has in Common with Anti-Pornography Laws, Campaign Finance Reform, and Telecommunications Regulation, 42 B.C. L. Rev. 1, 13 (2000); Dianne Leenheer Zimmerman, Information as Speech, Information as Goods: Some Thoughts on Marketplaces and the Bill of Rights, 33 Wm. & Mary L. Rev. 665, 709 (1992) (describing how many cases are determined by the discretion and sense of justice of each particular judge and jury due to the vagueness surrounding the scope of property rights and its intersection with speech rights).

\textsuperscript{90} \textit{See} Merges, Are You Making Fun of Me?, \textit{supra} note 40, at 312.

\textsuperscript{91} Tushnet, \textit{supra} note 89, at 13.

\textsuperscript{92} \textit{Id}.

\textsuperscript{93} \textit{Id}. at 20–21; \textit{see} Zimmerman, \textit{supra} note 89, at 709.
cultural divides, such as race or gender. Where a work speaks to less obvious or well-accepted issues, such as intra-racial parodies or, in some cases, works connected to homosexuality, judges have tended to draw narrower lines around what constitutes a “comment.” Restricting the exception to critiques that are already well-accepted minimizes its value. Such a cautious standard threatens the ability of satire to illuminate unexamined tensions and only reinforces dominant ways of thinking. It also values overtly critical works over more subtle combinations of traditions.

In contrast to the economic school, more recent scholarship from cognitive and behavioral psychology suggests that understanding both the nature of the property subject and its social function is crucial to determining the proper allocation of property rights. Research demonstrates that status quo bias acts as a powerful barrier to private transactions involving resource entitlements. People tend to value resources more highly if such resources have been initially allocated to them than if those goods had been allocated somewhere else. Therefore, people are reluctant to part with resources that they own even if the transaction would put the resource to a more valuable use. Such research puts renewed emphasis on finding optimal default allocations to increase


95 Compare Campbell, 510 U.S. at 582–83 (finding a parody where a rap song indirectly criticized earlier romanticized urban setting of song by a white artist), with Parks v. Laface Records, 329 F.3d 437, 453–54 (6th Cir. 2003) (finding no parody under the Lanham Act where rap group used the name of civil rights icon Rosa Parks in a song obliquely about changing times and new black heroes). The divergent verdicts in Campbell and Parks v. Laface Records seem to rest more on discomfort with interpreting intra-racial commentary than actual differences between the “parodic messages” of the songs. Similarly, in the early nineties, judges were reluctant to allow recoding of popular symbols by groups with homosexual affiliation. See S.F. Arts & Athletics, Inc. v. U.S. Olympic Comm., 483 U.S. 522, 540–41 (1987) (forbidding use of label “Gay Olympic Games” for alternative sporting event despite widespread use of the term Olympic(s) by other organizations unaffiliated with the official games); MGM-Pathe Commc’ns, Co. v. Pink Panther Patrol, 774 F. Supp. 869, 877 (S.D.N.Y. 1991) (enjoining use of the Pink Panther symbol by anti-gay violence patrols).

96 For example, the popular and critically acclaimed Grey Album, which combined music from the Beatles’ White Album and Jay-Z’s Black Album, was withdrawn under legal challenge presumably because the work would not meet the legal definition of parody. Electronic Frontier Foundation, Grey Tuesday: A Quick Overview of the Legal Terrain, http://www.eff.org/IP/grey_tuesday.php (last visited Sept. 11, 2005).

97 See Lewinsohn-Zamir, supra note 21, at 1717–18, 1726 (reasoning objective value judgments about property entitlements, based on the quality and identity of property at issue, can create welfare-maximizing default rules); infra note 99.

social welfare and private utility since allocations are likely to be given legitimacy by recipients and so remain in place. Understanding how the resource at issue operates and is valued by different stakeholders enables the formulation of more effective default rules.  

B. Copyright Minimalists and the First Amendment Critique

Before turning to cognitive research, it is worth discussing the primary critique of the economic property-based account of secondary use. First Amendment scholars articulate a positive account of First Amendment values to support the interests of audience members in recoding popular works and fault the economic framework for failing to consider democratic values of free speech and autonomy. This critique effectively identifies problems with the prevailing regime. Since copyright restrictions generally trespass on First Amendment freedoms, however, the First Amendment alone offers little guidance in how to balance competing interests. Solutions rooted in free speech discount the importance of a right to exclude for investment in entertainment projects. Like the economic account, the free speech critique imperfectly reflects the dynamics of the information entitlements at issue.

The First Amendment critique focuses on the control that copyright gives to large media firms over development of and engagement with culture. With the advent of mass media, many of the symbols and narratives that permeate everyday life are commodified. Think here of the Star Wars movies, the Coca-Cola logo, and characters from popular television shows such as Bart and Lisa Simpson. People find fulfillment in using such conventional or widely understood symbols to express their commitment to cultural, political, or social groups.

99 Sunstein & Thaler, supra note 21, at 1162 (arguing that because default rules inevitably shape preferences, legal frameworks should attempt to guide consumers and individuals toward welfare-maximizing choices); cf. Russell Korobkin, The Status Quo Bias and Contract Default Rules, 83 CORNELL L. REV. 608, 674–75 (1998) (arguing that the status quo bias demonstrates the desirability of tailoring default rules to specific circumstances).


102 John Philip Jones & Jan Slater, What’s In A Name?: Advertising and the Concept of Brands 217–25 (2003) (describing Coca Cola brand loyalists who fanatically
They comment on their surroundings through creative mixing of traditions and reworkings of popular culture.\(^{103}\) Even the decision to copy verbatim someone else’s speech may serve democratic values. The assertion that someone’s words are relevant to a new situation is a particular political stance.\(^{104}\) Copyright law restricts “the social flow of texts, photographs, music and . . . other symbolic works,” a form of control that may deprive us of the optimal cultural conditions for democratic debate and expressive innovation.\(^{105}\)

A similar critique may be found among scholars who view the Internet and the new “networked society” as particularly fertile ground to encourage active participation in culture and self-government.\(^{106}\) The ease of digital transmission and reproduction enables users to transform common symbols and texts for their own purposes, and so “speak back” to more powerful entities.\(^{107}\) Networked communication raises the potential for collaboration and decentralized production and thus active engagement in the cultural sphere.\(^{108}\) This same digital technology, however, also enables greater tracking, surveillance, and blocking of expressive consumption.\(^{109}\) Use of expressive works leaves a trail on the Internet. Owners seek to charge fees or enjoin unauthorized viewing and distribution of works or derivatives. In this way, strong intellectual property rights threaten to undermine the Internet’s potential to foster a true democratic culture.

To counteract these trends, scholars have put forward positive theories of the First Amendment that would limit the amount of protection copyright owners may claim. For example, Neil Netanel has argued that inherent in First Amendment values is a need for offer-


\(^{104}\) Tushnet, supra note 89, at 17.

\(^{105}\) Coombe, supra note 101, at 1866.


\(^{108}\) See Balkin, supra note 106, at 11–13, 33–34.

\(^{109}\) See id. at 18.
ings from “diverse and antagonistic” sources. Yochai Benkler has advanced a similar thesis, arguing that because the First Amendment requires a robust public domain, it limits the rights in information that may be granted to private entities. Jed Rubenfeld argues that the First Amendment cabins copyright by protecting an absolute “freedom of imagination.” No person should require permission to exercise her imagination, nor should anyone be prevented from disseminating works of her own imagination. Rubenfeld thus takes issue not with copyright’s prohibition on reproduction, but on the derivative works right that prevents unauthorized users from reimagining cultural works.

These theorists agree that a positive theory of the First Amendment would require a broadening of copyright’s treatment of transformative uses beyond the narrow allowance for targeted parody. According to the free speech model, where a secondary work is “transformative,” that is, the author has added her own original expression along with the previous work to change its meaning or purpose, courts should not enjoin that speech. As with other prior restraints, the injunction of transformative works, which by definition contain original expression, raises a specter of government-assisted censorship. Such restrictions risk suppressing valuable thoughts and ideas that would lose force if expressed another way. As to claims that unfettered secondary use will erode audience demand, minimal-

---


113 *Id.* at 4.

114 *Id.* at 5.


ists see such concerns as improperly safeguarding the moral value or reputation of the work against criticism and comment.\footnote{Rubenfeld, \textit{supra} note 112, at 53–54; Tushnet, \textit{supra} note 103, at 675.}

One problem is that the First Amendment, by itself, does not direct how to choose between speakers.\footnote{See David McGowan, \textit{Why the First Amendment Cannot Dictate Copyright Policy}, 65 U. Pitt. L. REV. 281, 284–85, 304–05 (2004).} Some transformative uses are still substitutive, such as the adaptation of a novel into a screenplay. The screenplay author may add pages of new dialogue and invent new characters; this is often the case even for authorized adaptations.\footnote{See Goldstein, \textit{supra} note 65, at 220–21.} The author may use quite a bit of imagination. But once that adaptation is commercially released, it will satisfy much of the audience’s demand for a film version, thereby siphoning off a large chunk of the owner’s expected return.\footnote{For some commentators, this sort of piracy is worse than the literal kind because the secondary user appropriates a new, untapped market, whereas if the infringing work were limited to the same medium, it would only attract the same people who already had the opportunity to purchase the plaintiff’s work. \textit{Nimmer \& Nimmer, supra} note 4, § 13.03[E][2] at 13–92; \textit{cf. Goldstein, supra} note 47, at 7 (asserting that publishers will invest more in purchasing, marketing, and producing a work when they know they can control all derivative uses).}

Rubenfeld and Netanel would answer this dilemma by imposing a compulsory license or profit apportionment scheme whereby a judge or neutral tribunal calculates what portion of the second author’s return is due to the appeal of the first creator’s work.\footnote{See Netanel, \textit{supra} note 41, at 380–81; Rubenfeld, \textit{supra} note 112, at 57–58.} Such a system would separate the right to receive compensation for use from the right to control what form the use takes. Assuming proper management, a panel of experts potentially could arrive at a reasonable licensing fee. Furthermore, the shadow of government oversight would likely encourage private bargaining.

However appealing, this system preserves the uncertainty of the current doctrine and undermines copyright’s incentive functions. It preserves uncertainty because it entrusts judges to separate changes that involve imagination from changes that are introduced only to evade the reproduction right. This latter class would still be subject to injunction.\footnote{Rubenfeld, \textit{supra} note 112, at 55.} This inquiry is arguably as subjective as asking if a new work “comments” on an original.\footnote{See Michael Abramowicz, \textit{A Theory of the Derivative Right and Related Doctrines} 42–44 (2005) (unpublished paper, on file with the author).}
The regime also undermines incentives because it does not address overuse. Economics scholars argue that mass promotion of a popular character may saturate the public’s demand for that character, especially if, as seems likely, many people will try to reimagine the character into their own products at the same time.\textsuperscript{126} While there may be reasons to doubt that such effects are universal or widespread, especially for well-known works, it is not clear that negative externalities should be discounted altogether.\textsuperscript{127} As discussed further in Part II, empirical data does suggest that in certain circumstances overexposure to expressive works can decrease audience demand by altering the ability to recognize, understand or appreciate authorized versions.\textsuperscript{128} While this is unlikely to deter amateur creators and low-intensity investments in creative production, the risks of instant copycats and quick market saturation may discourage sustained investment in expressive works.\textsuperscript{129} It is hard to see how such an environment leaves audiences or owners better off. We would have solved the prob-

\textsuperscript{126} See supra notes 54–60 and accompanying text.

\textsuperscript{127} See Lemley, \textit{Ex Ante Versus Ex Post}, supra note 11, at 145–46. Professor Mark A. Lemley has argued that copyright owners do not need a right to exclude secondary users because the problem of overgrazing, or congestion externalities, is self-limiting. He gives four reasons for this claim. First, only a subset of works that are widely known across a culture are likely to be “overgrazed.” \textit{Id.} at 145. Second, if consumers prefer the original context of the work, demand will remain high for that version. \textit{Id.} at 146. If consumers prefer the rewrite, then we should examine our assumption that consumers prefer static meaning in their cultural symbols. \textit{Id.} Third, allowing competition to produce sequels and rewrites may spur the original owner to produce a newer version faster. \textit{Id.} Finally, the social value of allowing subversion of icons outweighs the social good of protecting them. \textit{Id.} at 145–46. Although compelling on many counts, this account of overgrazing is “ex post.” It assumes a world in which cultural icons already exist. Once control over many forms of reproduction and reuse is removed, it may be much harder to establish works with iconic, consensus meanings across a culture. \textit{Cf.} Hughes, \textit{supra} note 7, at 943, 960 (noting that in the absence of intellectual property rights, use of a symbol such as Madonna’s image could mean something entirely different to someone in Oregon than it would in Venice Beach, and eventually could become meaningless altogether). One should not go too far here though. Complementary regimes such as trademark and unfair competition may prevent some free-riding aside from copyright protection. Goldstein, \textit{supra} note 65, at 220; F. Scott Kieff, \textit{The Case Against Copyright: A Comparative Institutional Analysis of Intellectual Property Regimes} 7 (Stanford Law Sch. John M. Olin Program in Law and Econ., Working Paper No. 297, 2004), available at http://ssrn.com/abstract=600802. \textit{But see} Dastar Corp. v. Twentieth Century Fox Film Corp., 539 U.S. 23, 37–38 (2003) (holding that the Lanham Act does not prevent unaccredited copying of copyrighted work in which the copyright has lapsed).

\textsuperscript{128} See infra notes 228–32, 346–53 and accompanying text.

\textsuperscript{129} \textit{Cf.} Abramowicz, \textit{supra} note 125, at 28 (arguing that the purpose of derivative work right is to prevent rent dissipation).
lem of copyright cannibalizing free speech rights by allowing free speech to cannibalize copyright.\(^{130}\)

As Mark Lemley has argued, what is needed is an account of incentive and access issues that is specific to copyright law.\(^{131}\) Economic property analysis and free speech doctrine are useful here, but both neglect key aspects of how consumers actually choose and make use of expressive works. The economic account neglects the importance of communication and discussion about competing interpretations of symbolic works. The minimalists ignore significant audience interests in the ability to recognize and identify with cultural works. By paying attention to research regarding how we understand and consume information, we can arrive at rules that more precisely balance free speech and democracy interests with the stability and incentive functions underlying audience choice.

II. An Information-Processing Account of Secondary Use

Three arguments support using cognitive research to analyze secondary use of informational goods. First, from a historical perspective, copyright has ceased to subsist in rights to copy tangible goods and now protects against a much wider range of activity designed to evoke proprietary works in the minds of an audience. Beginning in the early 1900s, a dramatic increase in new technologies led Congress and the courts to define copyright broadly across formats by reference to the perceptual effect of uses. The precise boundaries of such rights have never been satisfactorily explained. Second, owners sue and judges enjoin unauthorized secondary uses under copyright at least in part to protect against overuse and degradation of consumer associations with popular brands. Here, too, the extent to which copyright can and should protect against changes in consumer perceptions is unclear.

Third, for consumers, works of information exist simultaneously in tangible format ready for consumption, and in memory in need of discussion and elaboration for proper encoding of experience. This duality gives consumers interests on both sides of the debate. Secondary use of expressive works of information can crowd the external marketplace and distort the ability to engage productively with favor-

\(^{130}\) Furthermore, copyright cannot be returned to its original status as a right of reproduction only. See Rubenfeld, supra note 112, at 55. New methods of distribution have inexorably altered how people consume expressive works. See infra notes 148–83 and accompanying text. A right of reproduction today would provide less protection than it did in 1856 because so many ways exist to “copy” a work without literal reproduction.

\(^{131}\) Rubenfeld, supra note 112, at 44–45.
ite works. Once experienced, however, free discussion of the elements of shared cultural works facilitates categorization of knowledge, understanding of self, and development of cultural ties with peers. These dynamics, drawn from consumer psychology research, reveal a collision between the goals of owner and audience that should inform the secondary use debate.

A. From a Right of Reproduction to Right of Conceptual Representation

1. Early History

The development of the Copyright Act in the eighteenth century reveals the roots of the current debate regarding overgrazing and secondary use. To enable exploitation of works in formats like theatre, radio, and film, owners sought to protect against use beyond literal reproduction. Owners fought for a right to represent the essence of a work in any manner capable of being perceived. The boundaries of this right to represent meaning across formats intersected with traditional audience reuse and recoding rights in a way that has not yet been satisfactorily resolved.

The original copyright acts in both Britain and the United States focused on the economics of the printing press, and so defined narrowly the exclusive rights granted to an author. The early British Copyright statute, the Statute of Anne, protected only the mechanical rights to “print, reprint or import” a book. The first U.S. Copyright statute was similarly limited to giving authors of maps and books the exclusive right to “print[], reprint[], publish[], and vend” such works. Courts in both jurisdictions read these rights to prohibit only literal copying, and allowed subsequent authors free reign to use existing works as raw material in abridgements and translations.

The creation of a property right in something as ethereal as literary doctrine troubled jurists from the beginning. In Millar v. Taylor, an early British case addressing copyright’s scope, one of the presiding

132 8 Ann., c. 19 (1710) (Eng.).
133 Act of May 31, 1790, ch. 15, 1 Stat. 124.
134 See, e.g. Folsom v. Marsh, 9 F. Cas. 342, 343 (C.C.D. Mass. 1841) (No. 4901) (distinguishing between “facile use of . . . scissors” to condense a work and steal its most valuable portions with “real, substantial condensation,” which revealed “intellectual labor and judgment” of bona fide new author); Burnett v. Chetwood, (1720) 35 Eng. Rep. 1008, 1009–10 (Ch.) (reasoning that the statute restrains only “the mechanical art of printing” and not new translations).
judges questioned the boundaries of this new form of intangible property:

[T]he property here claimed is all ideal; a set of ideas which have no bounds or marks whatever . . . . Their whole existence is in the mind alone; incapable of any other modes of acquisition or enjoyment, than by mental possession or apprehension . . . .

In response to similar arguments in an earlier case, William Blackstone articulated what was to become the standard response: that the property right was not in the ideas put forth in the book but was found in its actual text. Blackstone elaborated on this description in his second volume of the Commentaries published a few years later. Blackstone wrote that “[t]he identity of a literary composition consists entirely in the sentiment and language, the same conceptions, cloaked in the same words, must necessarily be the same composition . . . .” He classified literary property as a type of estate similar to the natural right of occupancy in unclaimed land.

The limit of copyright to literal copying showed concern for the rights of secondary authors but also sought to make expressive works available to the widest possible audience. Abridgements served a valuable public interest by making more expensive, and in some cases more complicated, works available in cheaper and simpler versions. By protecting the ability of secondary authors to make use of previous works, judges believed they were furthering aims of public education and intellectual progress. That an unauthorized derivative work

137 William Blackstone, 2 Commentaries *406; see Mark Rose, Authors and Owners: The Invention of Copyright 89–90 (1993).
139 Id. at *400–01, 405; Rose, supra note 137, at 90.
140 See Banker v. Caldwell, 3 Minn. 94, 3 Gil. 46, 53 (1859) (“It has been a very difficult question in the courts to determine what is original and what borrowed or pirated, in a literary production. It cannot be necessary that the matter contained in a work, the thought, sentiment and language, should be all original, to entitle the author to the protection of a copyright, because if such was the case, in the present advanced state of the sciences, learning, and literature, we might look for very few additions which would fall in the privileged sphere. It would exclude . . . abridgments of works beyond the reach of many, which now form a large portion of the means through which knowledge is conveyed to the people.”).
might displace the market for the original or a licensed abridgement or translation was not given much weight.\textsuperscript{141}

Beyond exact text, the plot, characters, and sensibility of the author were fair game for imitators and critics. For example, in 1853 the author Harriet Beecher Stowe sued the translators of an unauthorized version of \textit{Uncle Tom's Cabin} targeted to German immigrants.\textsuperscript{142} The 1831 Copyright Act in force at the time was silent on the issue of derivative rights.\textsuperscript{143} Stowe argued that the translation was a direct copy, with only the mechanical signs having been changed.\textsuperscript{144} The judge disagreed and read the Act to protect only the precise words used by Stowe.\textsuperscript{145} He found that her characters, concepts, and creations were public property, however, and “may be used and abused by imitators, play-rights and poetasters.”\textsuperscript{146} After publication, “[t]he author’s conceptions have become the common property of his readers, who cannot be deprived of the use of them, nor of their right to communicate them to another clothed in their own language . . . .”\textsuperscript{147}

2. “Cognate” Rights

Soon after the \textit{Stowe} decision, Congress began to expand copyright’s scope to include non-literal copies.\textsuperscript{148} It is the boundaries of this expansion that have yet to be fully mapped out. For example, in 1856, Congress had added the right to perform a dramatic work publicly to the exclusive rights of copyright owners.\textsuperscript{149} \textit{Daly v. Palmer}, de-

\begin{footnotesize}
\begin{enumerate}
\item[141] In an 1847 case, Judge McLean of the First Circuit Court of Appeals criticizes the fair abridgement doctrine for its lack of attention to the effects on the sale of the work abridged. \textit{See} \textit{Story} v. \textit{Holcombe}, 23 F. Cas. 171, 172–73 (C.C.D. Ohio 1847) (No. 13,497) (“The argument that the abridgement is suited to a different class of readers, by its cheapness, and will be purchased on that account by persons unwilling and unable to purchase the work at large, is not satisfactory. This to some extent may be true; but are there not many who are able to buy the original work, that will be satisfied with the abridgement? . . . The reasoning on which the right to abridge is founded, therefore, seems to me to be false in fact. It does to some extent in all cases, and not unfrequently to a great extent, impair the rights of the author—a right secured by law.”). Finding himself bound by precedent based on the English doctrine, however, he refused to fully enjoin an abridgement of Justice Story's Commentaries on Equity Jurisprudence.
\item[143] \textit{See} Act of Feb. 3, 1831, ch. 16, 4 Stat. 436.
\item[144] \textit{Stowe}, 23 F. Cas. at 202–05.
\item[145] \textit{Id.} at 206–08.
\item[146] \textit{Id.} at 208.
\item[147] \textit{Id.} at 206.
\end{enumerate}
\end{footnotesize}
cided in 1868, defined the boundaries of an infringing performance.150 In that case, the plaintiff wrote a popular play involving a dramatic rescue in front of an oncoming train.151 A British playwright obtained a copy and wrote a different hit play in London containing a similar scene.152 A New York theatre sought to produce the second play and Daly sought an injunction.153 The two plays shared little in common in terms of dialogue, characters, or plot save for the infamous railroad scene, and even that contained material differences.154 The new right of public performance, however, freed the court from side-by-side comparison as a test for infringement.155 Instead, the court found that Daly’s play could be infringed “by representation as well as by printing, publishing, and vending.”156 The test of whether a dramatization violated this right of “representation” was its ability through movement and gesture to “excite[] emotions and impart[] impressions” that the audience would experience as substantially similar to the original work.157 Daly thus marked the first moment where the rights protected by copyright ceased to subsist entirely in tangible works and instead became focused on the impact of the work on paying audiences.158

Technological advances pushed the boundaries of prohibited “copying” further. Mechanical piano rolls first presented the question of whether non-literal musical reproduction constituted a “reproduction” prohibited under the Act.159 These rolls, perforated cylinders designed to produce melodies when rotated inside a player piano, gave the same impression of a person sitting and playing a composition off of sheet music. Music publishers sought royalties for such use, but the pianola manufacturers argued that a perforated cylinder that little resembled a piece of paper containing musical notations could not violate copyright’s reproduction right.160 The courts were thus

151 Id. at 1133.
152 Id. at 1133–34.
153 Id. at 1134.
154 See id. at 1134–35.
155 See Daly, 6 F. Cas. at 1135–38.
156 Id. at 1137–38.
157 Id. at 1137. The court found that “[a]ll that is substantial and material in the plaintiff’s ‘railroad scene’ has been used . . . in the same order and sequence of events, and in a manner to convey the same sensations and impressions to those who see it represented, as in the plaintiff’s play.” Id. at 1138.
158 See Goldstein, supra note 65, at 213–14.
159 Goldstein, supra note 47, at 65.
160 Id. at 66.
presented with the question of whether a piano roll was a “copy” of sheet music.\footnote{White-Smith Music Publ’g Co. v. Apollo Co., 209 U.S. 1, 8–11 (1908).}

By the time the music rolls case reached the Supreme Court in 1908, in \textit{White-Smith Music Publishing Co. v. Apollo Co.}, the manufacture and sale of such pianos had reached seventy five thousand and the sale of the rolls had topped over one million per year.\footnote{\textit{Id.} at 9.} Accordingly, the Court chose a narrow reading of the then-current act and limited the definition of “copy” to “reproduction or duplication of a thing.”\footnote{\textit{Id.} at 17.} “The statute has not provided for the protection of the intellectual conception apart from the thing produced,” wrote Justice Day for the majority.\footnote{\textit{Id.}} Justice Holmes concurred, but wrote separately to argue that to protect inventors fully, the statute should be revised to extend protection not only to the form of a composition, but to its “essence.”\footnote{\textit{Id.} at 19 (Holmes, J., concurring).} “A musical composition is a rational collocation of sounds apart from concepts . . . . On principle anything that mechanically reproduces the sounds ought to be held a copy . . . .”\footnote{\textit{Apollo}, 29 U.S. at 19–20 (Holmes, J., concurring).}

The publishers then took the matter up with Congress, which was at the time in the process of holding hearings to revise the copyright laws.\footnote{Goldstein, \textit{supra} note 47, at 67; see also Metro-Goldwyn-Mayer Distrib. Corp. v. Bijou Theatre Co., 59 F.2d 70, 74 (1st Cir. 1932) (describing the desire to overturn the \textit{Apollo} decision as one of the purposes of the 1909 Copyright Act).} In 1909, Congress passed the 1909 Copyright Act, which added several new rights for copyright holders.\footnote{Act of March 4, 1909, ch. 320, § 1(d), 35 Stat. 1075 (current version at 17 U.S.C. § 106 (2000)).} In the case of dramatic works, these included the right to “represent” the work in whole or in part “in any manner or by any method whatsoever”\footnote{\textit{Id.}} and, in the case of musical compositions, the right to “make any arrangement or setting of it or of the melody of it in any system of notation or any form of record in which the thought of an author may be recorded and from which it may be read or reproduced . . . .”\footnote{\textit{Id.} § 1(e). The Act made this latter right only prospective and provided a compulsory license to ensure that no producer could gain a monopoly over the entire format. See Goldstein, \textit{supra} note 47, at 67.} These rights to “represent” intangible works or reproduce “thought” quietly but ex-
plicitly expanded copyright’s reproduction right from tangible objects into the realm of perceptual similarities.

This transformation was complete when Holmes decided a 1911 case, *Kalem Co. v. Harper Bros.*, concerning a motion picture photoplay.\(^{171}\) The case, involving a silent movie based on the novel *Ben-Hur*, turned on the right added in 1870 to “dramatize” a book.\(^{172}\) Although the silent movie was not literally similar to the text of the book, Holmes relied on *Daly* to find that “[t]he essence of the matter . . . is not the mechanism employed but that we see the event or the story lived.”\(^{173}\)

Holmes’ insistence on locating the “essence” of the protected work recalls Blackstone’s attempts to identify what was “essential” in copyright. Where Blackstone had located the borders of the literary estate at the “sentiment and language . . . the same conceptions cloathed in the same words,”\(^{174}\) Holmes had by 1911 extended the boundaries to include anything that conveyed the same story to the understanding of the audience. The 1909 Copyright Act and Holmes’ *Kalem* opinion paved the way for courts to apply the right of conceptual representation embraced in *Daly* to new technologies from the printing press such as moving pictures, radio, television, and even computer software.\(^{175}\) Any method of production that might call up the “essence” of the composition in the mind of a paying audience now fell within the limits of copyright.

The current Copyright Act reflects this expansion of the author’s domain. In the 1976 Copyright Act, the separate sections relating to dramatic, literary, musical, or photographic works were for the most part combined in favor of a unitary approach to the bundle of rights granted under copyright.\(^{176}\) The exclusive reproduction right granted to owners includes the right to produce a material object in which the work can be “perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device.”\(^{177}\)

\(^{171}\) See 222 U.S. 55, 61–62 (1911).

\(^{172}\) See id. at 61.

\(^{173}\) Id.

\(^{174}\) Blackstone, supra note 137, at *406.

\(^{175}\) See generally Buck v. Jewell-LaSalle Realty Co., 283 U.S. 191 (1931) (radio); Apple Computer, Inc. v. Franklin Computer Corp., 714 F.2d 1240, 1250–54 (3d Cir. 1983) (computer operating systems); *Bijou Theatre Co.*, 59 F.2d at 75 (movies).


\(^{177}\) Id. §§ 101, 102.
tive work right reserves to the owner "any other form in which a work may be recast, transformed, or adapted."\footnote{Id. §§ 101, 106.}

Though this language has existed for over twenty years, courts have not carefully examined the outer limits of this right of conceptual representation.\footnote{See Nimmer & Nimmer, supra note 4, § 13.03[A] at 13-35 to -36 (noting that the distinction between literal and non-literalt similarity in copyright has received "almost no express judicial recognition").} The "substantial similarity" test, which determines the boundaries between infringing and non-infringing similar works, offers little guidance.\footnote{Judge Learned Hand, the architect of the foundational approach to the modern substantial similarity inquiry in Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. 1930), concluded in one of his last copyright cases that "[o]bviously, no principle can be stated as to when an imitator has gone beyond copying the 'idea' and has borrowed its 'expression.' Decisions must therefore inevitably be ad hoc." Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960).} Most courts look to an "ordinary observer" or "audience" test that asks whether the new work evokes a similar "impression" as the old,\footnote{See Incredible Techs., Inc. v. Virtual Techs., Inc., 400 F.3d 1007, 1011 (7th Cir. 2005) (describing the "ordinary observer" test in the context of video games as "whether the accused work is so similar to the plaintiff's work that an ordinary reasonable person would conclude that the defendant unlawfully appropriated the plaintiff's protectible expression by taking material of substance and value"); Stromback v. New Line Cinema, 384 F.3d 283, 297 (6th Cir. 2004) (same, and concluding that the question turns on whether the "net impression" of the works' expressive elements are substantially similar to one another); Sturdevia v. United Arab Emirates, 281 F.3d 1287, 1296 (D.C. Cir. 2002) (same); Yankee Candle Co., Inc. v. Bridgewater Candle Co., 259 F.3d 25, 33 (1st Cir. 2001) (same); Educ. Testing Servs. v. Katzman, 793 F.2d 533, 541 (3d Cir. 1986) (same); see also Yurman Design, Inc. v. PAJ, Inc., 262 F.3d 101, 111 (2d Cir. 2001) (describing the ordinary observer test as whether "an average lay observer would recognize the alleged copy as having been appropriated from the copyrighted work").} or whether an ordinary observer would conclude that the new work incorporated something of value from plaintiff's work.\footnote{See Nimmer & Nimmer, supra note 4, §§ 13.03[E][1] at 13–82, 13.03[E][2] at 13-89 to -90.\footnote{See Nimmer & Nimmer, supra note 4, § 13.03[A][1][c] at 13-46.} These tests expand the limits of copyright beyond literal similarity or even semantic similarity to any work that, in the opinion of the average viewer or listener, evokes a former work. Such tests assume without explanation that plaintiffs deserve control over any aspect of their work with value to the audience, and at the extreme threaten to elide copyright's cornerstone distinction between protected expression and unprotectible ideas and concepts.\footnote{See Nimmer & Nimmer, supra note 4, § 13.03[A][1][c] at 13-46.}
3. Effect on Common Law Traditions of Parody, Burlesque, and Pastiche

The confusion over copyright’s boundaries is especially pernicious in cases where a new author seeks to reuse excerpts of an existing work for a different purpose. To date, no court has clearly defined in the wake of the 1909 expansions whether the ability to communicate subjective conceptions of an expressive work belongs to the audience or to the author.\(^\text{184}\) Instead, resolution of this question has come to focus on imprecise questions about the secondary user’s motivation in using the work.

Although nothing in the legislative history indicates that Congress intended the 1909 Copyright Act to alter the common-law rights of audience members to parody and burlesque popular works, the new derivative rights did just that.\(^\text{185}\) Before the revisions of 1909, the law took a benign view of parody, satire, burlesque, pastiche, and other forms of communicating alternative audience perceptions of a cultural work. Such use was common as a form of commercial art and in everyday discourse. On the stage, most popular plays were the subject of parodies that were performed at about the same time as the originals.\(^\text{186}\) Imitators often spoofed the styles of well-known writers and performers. Newspapers reported or reproduced parodies of popular works fre-

---

\(^{184}\) Compare Warner Bros., Inc. v. Am. Broad. Cos., Inc., 720 F.2d 231, 242 (2d Cir. 1983) ("Stirring one’s memory of a copyrighted character is not the same as appearing to be substantially similar to that character, and only the latter is infringement."). \(^{185}\) S. Rep. No. 6187, at 1 (1907) (stating that the bill preserved common law adjudication of what constituted infringement and fair use); see also H.R. Rep. No. 2222, at 7, 9 (1909) (stating that monopoly given to the author was ultimately intended to benefit the public and that nothing in the bill intended to alter author’s common-law rights with respect to the work).

\(^{186}\) Yankwich, supra note 83, at 1136; William Lyon Phelps, \textit{Daily Talk About Books and Authors}, Wash. Post, Dec. 12, 1933, at 6 (noting how in past years two New York comedians always selected the most popular play of the year to burlesque).
quently.\textsuperscript{187} As far as courts were concerned, so long as the parodist had not copied the previous work in its entirety, or set out to replace rather than ridicule the original, parodies qualified as fair use.\textsuperscript{188}

Not long after passage of the 1909 Copyright Act, judicial attitudes towards parody and burlesque took a marked shift. The ability to exploit a single work through the different mediums of radio, motion picture, and television led owners of popular dramatic works to clamp down on parodists who transposed the work to a new format. Judges began vilifying the act of parody itself as a taking no different from any other theft.\textsuperscript{189}

For example, in 1955, a California district court enjoined a television parody of a popular film.\textsuperscript{190} The decision in the case, \textit{Loew’s, Inc. v. Columbia Broadcasting System}, explicitly distinguished between acceptable and unacceptable formats for parody.\textsuperscript{191} It drew a line between literature and scientific texts, where broad fair use and parody might be allowed because such works aimed to further progress in the arts and sciences, and fields where “business competition” existed such as broadcast entertainment.\textsuperscript{192} The court ruled that the factor of whether “the infringing work . . . has been issued for commercial gain, rather than in the interests of advancement of learning” was cru-

\textsuperscript{187} See \textit{A Parody}, N.Y. Times, Aug. 23, 1896, at 2 (reprinting a parody of Horace by Alexander Pope); \textit{A Parody}, N.Y. Times, May 3, 1896, at 11 (reprinting an anonymous parody of Emerson’s Brahma); \textit{Barrie Denies Parody}, N.Y. Times, July 17, 1921, at 23 (reporting the denial of Sir James Barrie that he had written a parody of a Mrs. Asquith’s memoirs reportedly entitled “Knees I Have Sat Upon.”); Albert E. McKay, \textit{Parodies on “The White Man’s Burden.”}, N.Y. Times, May 27, 1899, at BR351 (reprinting portions of a number of parodies of Kipling’s poem); \textit{Parody}, N.Y. Times, Mar. 29, 1885, at 12 (essay praising parodies for their role in chastening and instructing authors).

\textsuperscript{188} See generally Green v. Luby, 177 F. 287 (C.C.S.D.N.Y. 1909) (singing an entire song to imitate popular singer took more than necessary and was infringement); Green v. Minzenheimer, 177 F. 286 (C.C.S.D.N.Y. 1909) (singing the chorus and one verse of a copyrighted song to parody the style of a popular singer was not infringement); Bloom & Hamlin v. Nixon, 125 F. 977 (C.C.E.D. Pa. 1903) (same with respect to the chorus only).

\textsuperscript{189} See \textit{Loew’s, Inc. v. Columbia Broad. Sys., Inc.}, 131 F. Supp. 165, 183 (S.D. Cal. 1955), \textit{aff’d sub nom.} Benny v. Loew’s Inc., 239 F.2d 532 (9th Cir. 1956), \textit{aff’d sub nom. by an equally divided court}, Columbia Broad. Sys., Inc. v. Loew’s, Inc. 356 U.S. 43 (1958); Hill v. Whalen & Martell, 220 F. 359 (S.D.N.Y. 1914) (characters in play \textit{In Cartoonland} were clearly meant to be copyrighted Mutt and Jeff characters and so threatened to meet consumer demand for an authorized dramatization of the strip).

\textsuperscript{190} \textit{Loew’s}, 131 F. Supp. at 186.

\textsuperscript{191} \textit{Id.} at 175, 183.

\textsuperscript{192} \textit{Id.} at 175 (“[T]he law implies the consent of the copyright owner to a fair use of his publication for the advancement of the science or art . . . We do not think . . . [the use of the word art was used in a sense broad enough to include a T.V. program . . .]”) (emphasis added).
cial.\textsuperscript{193} Many commentators at the time, including Leon Yankwich, the chief judge of the district, condemned the decision for its cramped view of parody and satire and its contention that the author’s desire for commercial gain was “primary” in determining fair use.\textsuperscript{194} Nonetheless, the decision was upheld and praised by both the Ninth Circuit and the U.S. Supreme Court.\textsuperscript{195}

The \textit{Loew’s} case introduced the dichotomy, picked up by the Supreme Court later in the 1994 decision in \textit{Campbell v. Acuff-Rose Music, Inc.} between “true” parody that seeks to make a critical comment and false parody, which looks only for attention or commercial profit.\textsuperscript{196} Although the Court in \textit{Campbell} provided greater leeway for commercial parodies than in \textit{Loew’s}, it preserved the rhetorical distinction between true, disinterested critique and crass, attention-getting commercialism.\textsuperscript{197} The result has been a test that for the most part distinguishes between commercial and non-commercial, or private and public, expressive activity.\textsuperscript{198} Since \textit{Loew’s} and \textit{Campbell}, parody and pastiche have greatly retreated from the commercial marketplace. Those seeking to rework a popular text, comment on favorite characters through fiction, or call to mind iconic works in fine art, for example, risk onerous lawsuits.\textsuperscript{199}

Digital technology, however, destabilizes the tenuous balance drawn in cases like \textit{Loew’s} and \textit{Campbell}. The Internet is quickly destroying the boundaries between personal and public, and private and commercial in communication. Private use of works, once considered outside the scope of copyright, now takes place in increasingly public fora. For example, Marvel Comics recently brought a lawsuit to pre-

\textsuperscript{193} Id. at 184–85.

\textsuperscript{194} See Yankwich, \textit{supra} note 83, at 1151 (pointing out that “[m]aterial gain by writers through parody and burlesque is not of modern origin,” and that writers from the Greeks through the great English parodists of the 15th and 16th centuries received compensation for writing satires).

\textsuperscript{195} \textit{Benny}, 239 F.2d at 534, 537.

\textsuperscript{196} See \textit{Campbell v. Acuff-Rose Music}, 510 U.S. 569, 580–81 (1994); \textit{Loew’s}, 131 F. Supp. at 174–76; see also \textit{MCA, Inc. v. Wilson}, 677 F.2d 180, 182 (2d Cir. 1981) (“While commercial motivation and fair use can exist side by side, the court may consider whether the alleged infringing use was primarily for public benefit or for private commercial gain.”); \textit{Wainwright Secs., Inc. v. Wall Street Transcript Corp.}, 558 F.2d 91, 94 (2d Cir. 1977) (“[Fair use] distinguishes between a true scholar and a chiseler who infringes a work for personal profit.”).

\textsuperscript{197} See \textit{Campbell}, 510 U.S. at 578–85.

\textsuperscript{198} See \textit{Lessig, supra} note 13, at 8.

vent fans of its superhero comics from “dressing up” as their protected characters in an unaffiliated fantasy game. Marvel obviously wants to protect the audience for its authorized games, but its stance is not so distinct from attempting to prohibit pretend games in one’s backyard.

Distance-learning courses have similarly challenged the traditional fair use exception for educational discussions and performances. The original exception assumed a face-to-face interaction in a classroom setting. In 2002 Congress had to pass a special rule clarifying that the exception applied to online learning environments as well. For-profit institutions, which are increasingly prevalent, and libraries, which host many such courses, however, remain outside the exception. The ability of students in those classes to “discuss” copyrighted works online free of charge remains in doubt.

To say that copyright extends to audience perception then does not answer the question of where to draw copyright’s boundaries. Although the question arises in numerous fair use cases, courts have not clarified at what point an audience’s memory and perception of characters, music or other expressive elements becomes the property of the audience. The proper balance between the owner’s right to profit from the original creation and the audience’s right to identify with, consider, rework, and discuss the significance of the work remains unresolved.

For this reason, those who argue that the dynamics of digital networks should control the contours of user rights may aim too narrowly. These technologies are still developing and will change many times in the coming decades. What will remain the same, or at least will shift more slowly over time, are the mechanics of human comprehension. Just as owners have fought for rights that are purposefully not technology-specific but extend to anything that sends a cognitive message, so should user rights not depend on technology, but should


202 Hutchinson, supra note 201, at 2225–27.

203 See infra notes 271–321 and accompanying text.

204 See, e.g., Balkin, supra note 106, at 52.
be grounded in the workings of information perception. Otherwise, the battle that Congress feared for authors, in which they would have to litigate every new technological advance to determine protection, is being fought by users to defend a diminishing universe of fair uses with each new technological breakthrough.

B. Copyright Ownership in a Desired Brand Identity

The extension of copyright into amorphous realms of consumer perception has led authors and owners to claim an ownership right in the preservation of a desired brand identity. From their standpoint, the right to represent a story or song across mediums must include the ability to halt competing representations of the same work, even if the two works are not confusingly similar or do not serve the same purpose or audience. Because profitability in entertainment and increasingly retail industries depends on being able to repurpose familiar popular works across a variety of media, owners must vigilantly protect core works to ensure that varied incarnations maintain a consistent meaning and identity to consumers. The introduction of negative or dissonant associations with a marquee work threatens the ability of that work to support large tie-in and promotional enterprises. This, at least, is the harm claimed when owners sue to enjoin unauthorized secondary use. By adding a copyright claim to what are essentially trademark concerns, owners are able to sidestep trademark’s broader free speech safeguards. Courts have been inconsistent in their willingness to entertain claims based on alteration in consumer perception in the context of copyright. The uncertainty of both trademark and copyright rules in this area has left consumers and secondary users without clear guidance.

1. Positive and Negative Associations

Due to their negligible marginal cost, entertainment companies increase profits by producing and distributing products to as many audiences and markets as possible. Producers may distribute

---

206 See, e.g., Michael J. Wolf, The Entertainment Economy 25 (1999) (noting how the need to attract audience attention has led studios to consolidate assets with networks, cable TV stations, and book publishers to market and distribute products through a variety of outlets).
207 See id. at 97–98, 228–30.
profitable works through new channels, such as releasing films on television and DVDs. Many films earn more through DVD sales than they do in their initial box office run. Marketers repurpose and recycle profitable works through sequels, spin-offs, adaptations, and promotional tie-ins. Books become movies, movies reappear as musicals or video games, and popular songs animate advertising or films. Once a work has achieved commercial success, its audience will follow it from product to product. Tie-in products increase loyalty by heightening familiarity and engagement with the work. Some brands earn more through merchandising than through the original product. Winnie-the-Pooh, for example, has grown from a tubby little bear to a billion dollar franchise for Disney through sales of DVDs, toys, and his popular theme-park ride, the Honey Hunt.

The ability to repurpose characters and brand images lowers risks and costs for owners. Use of an already popular story line or character offers more predictable revenues and requires less promotion. No matter what, the tie-in product will interest some segment of the audience. The strength of established works is not just theory. Of the top ten films of all time by revenue, for example, five are sequels or prequels, and four are based on popular books or comics.

---


210 Wolf, *supra* note 206, at 231 (“When a company offers theme parks, hotels . . . movies, toys, fast food, books, videos, records, magazines, clothing, and other products, the hope is that all these products and the efforts behind them will mesh and contribute to a chain reaction that creates more energy, awareness and economic effect than any single aspect might have done on its own.”); see John O’Shaughnessy & Nicholas Jackson O’Shaughnessy, *Persuasion in Advertising* 63 (2004); Pessach, *supra* note 78, at 1093.

211 See Laura Bradford, *Who Owns Pooh?: Disney Fights in Court over Digital Rights to a “Silly Old Bear” as New Technology Casts Doubt on Old Contracts*, Time, Jul. 15, 2002, at B16. Another example is the *Jurassic Park* movie series. The first two movies themselves garnered $1.5 billion in worldwide box office sales, while receipts from home videos, toys, video games, amusement park rides, and other revenue streams has approached $3.5 billion. Wolf, *supra* note 206, at 229.


213 Wolf, *supra* note 206, at 224–25 (stating that we are predisposed to accept, or at least look at, anything that carries the brand of a big entertainment company).

214 All Time Box Office Domestic Grosses, http://www.boxofficemojo.com/alltime/domestic.htm (last visited Sept. 11, 2005). The five sequels are *Shrek 2*, *Lord of the Rings: Return of the King*, *Star Wars I: The Phantom Menace*, *Spider-Man 2*, and *Star Wars III: Revenge of the Sith*. In addition to *Spider-Man 2* and *Lord of the Rings*, the other films based on books
Identifiable brands form the center of this juggernaut. For entertainment products, characters, situations, melodies, authors, and designs can all act as brand signifiers.\textsuperscript{215} To the purchasing public, the use of a popular character or recognizable tune acts as an information shortcut or heuristic. Consumers save time and energy evaluating a product by conferring on the product their associations with the brand.\textsuperscript{216} To perform this function effectively, brands must maintain a consistent personality and identity.\textsuperscript{217} In fact, research studies show that consumers punish products and brands that require too much effort to evaluate.\textsuperscript{218} A brand that stands for something simple is “more memorable, more visible and more meaningful” than a brand associated with a variety of images.\textsuperscript{219} To keep parents and children invested in purchasing new Winnie-the-Pooh DVDs, for example, Disney must ensure that he maintains a loveable and innocent persona through all licensed uses.

Marketers know that whatever is associated with a brand has the power to affect its image. Mechanisms in our brains automatically classify things as “good” or “bad” as soon as we perceive them.\textsuperscript{220} If something that anchors an entertainment brand, such as a popular character or a well-known song, is associated with something that triggers a positive emotional response, our appreciation of the brand increases.\textsuperscript{221} Accordingly, much of modern advertising is designed to

\textsuperscript{215} See O’Shaughnessy & O’Shaughnessy, supra note 210, at 40–41, 60; see also Cristel A. Russell et al., People and “Their” Television Shows: An Overview of Television Connectedness, in THE PSYCHOLOGY OF ENTERTAINMENT MEDIA: BLURRING THE LINES BETWEEN ENTERTAINMENT AND PERSUASION 275, 284 (L.J. Shrum ed., 2004) (arguing that television shows act as brands).

\textsuperscript{216} O’Shaughnessy & O’Shaughnessy, supra note 210, at 60; see David N. Martin, Romancing the Brand: The Power of Advertising and How to Use It’ 89 (1989); Wolf, supra note 206, at 223.

\textsuperscript{217} Martin, supra note 216, at 89–92; O’Shaughnessy & O’Shaughnessy, supra note 210, at 66; see Wolf, supra note 206, at 223–24.

\textsuperscript{218} Ellen C. Garbarino & Julie A. Edell, Cognitive Effort, Affect, and Choice, 24 J. CONSUMER RES. 147, 156 (1997).

\textsuperscript{219} O’Shaughnessy & O’Shaughnessy, supra note 210, at 66; see Martin, supra note 216, at 92; Wolf, supra note 206, at 223.


increase positive associations with advertised brands and products through use of symbolic and emotional appeals to the values of the target audience.\textsuperscript{222}

Negative information, however, has an even greater impact on brand association than positive information.\textsuperscript{223} Brand images are developed over time.\textsuperscript{224} Consumers learn to “trust” a brand through repeated exposure.\textsuperscript{225} If a brand somehow has been associated with incompatible values or unpleasant images, consumers will be less likely to purchase it.\textsuperscript{226} Studies suggest that the negative perception of the brand will persist for some time even through attempts by the owners to provide countering information.\textsuperscript{227}

Frequent exposure can also harm consumer attitudes. During initial introduction to a brand or message, frequency of exposure increases familiarity and likeability.\textsuperscript{228} Individuals tend to misattribute the increased ease of processing of a familiar message to the content of the message itself, and so find their tolerance of and positive attitude toward the message increase over time.\textsuperscript{229} Consumers thus will

\begin{itemize}
  \item increase positive associations with advertised brands and products through use of symbolic and emotional appeals to the values of the target audience.\textsuperscript{222}
  \item Negative information, however, has an even greater impact on brand association than positive information.\textsuperscript{223} Brand images are developed over time.\textsuperscript{224} Consumers learn to “trust” a brand through repeated exposure.\textsuperscript{225} If a brand somehow has been associated with incompatible values or unpleasant images, consumers will be less likely to purchase it.\textsuperscript{226} Studies suggest that the negative perception of the brand will persist for some time even through attempts by the owners to provide countering information.\textsuperscript{227}
  \item Frequent exposure can also harm consumer attitudes. During initial introduction to a brand or message, frequency of exposure increases familiarity and likeability.\textsuperscript{228} Individuals tend to misattribute the increased ease of processing of a familiar message to the content of the message itself, and so find their tolerance of and positive attitude toward the message increase over time.\textsuperscript{229} Consumers thus will
\end{itemize}

\textsuperscript{222} See O’Shaughnessy & O’Shaughnessy, \textit{supra} note 210, at 57–59, 64; see also Martin, \textit{supra} note 216, at 95 (“In advertising, we seek to shape attitudes . . . . What we want is for the consumer to remember the brand and what it offers and to have a positive attitude about trying it.”).

\textsuperscript{223} O’Shaughnessy & O’Shaughnessy, \textit{supra} note 210, at 63; see also Stephen J. Hoch & Young-Won Ha, \textit{Consumer Learning: Advertising and the Ambiguity of Product Experience}, 13 J. Consumer Res. 221, 223 (1986) (stating that because of asymmetric costs of different types of errors, people are more wary of mistakenly accepting something that might be bad than mistakenly rejecting something that might be good).

\textsuperscript{224} Martin, \textit{supra} note 216, at 89.

\textsuperscript{225} O’Shaughnessy & O’Shaughnessy, \textit{supra} note 210, at 63, 67.

\textsuperscript{226} See O’Shaughnessy & O’Shaughnessy, \textit{supra} note 210, at 64–65; cf. Jennifer Aaker et al., \textit{When Good Brands Do Bad}, 31 J. Consumer Res. 1, 13 (2004) (summarizing study that showed that consumer relationships with “sincere” brands, defined as those like Hallmark or Ford that show traits of nurturance, warmth, family-orientation, and traditionalism, suffered more after “transgression” by the brand, than consumer relationships with exciting, “trendy” brands, such as MTV or Virgin).

\textsuperscript{227} O’Shaughnessy & O’Shaughnessy, \textit{supra} note 210, at 59–60.

\textsuperscript{228} See Robert B. Zajonc & Hazel Marcus, \textit{Affective and Cognitive Factors in Preferences}, 9 J. Consumer Res. 123, 125 (1982).

respond more favorably to messages communicated frequently even if they initially disagree with the message or find its source not credible. After a certain level of unvaried exposure, however, tedium sets in and consumers will begin to feel negatively toward the message or brand. Owners of expressive properties carefully manage audience exposure to enhance familiarity but to decrease the likelihood of audience fatigue.

These marketing truisms explain the panic of brand managers when an outsider uses a brand element in an unauthorized way. Although owners may eschew smaller players when it comes to licensing, they will pay attention to even slight secondary uses done without permission. Such uses threaten to taint carefully managed brand campaigns by association with negative or even just inconsistent elements.

According to media reports and publicly filed complaints, this harm to the perceptive value of a character or work is precisely the concern of litigants seeking to enjoin secondary uses. For example, after a flash animation film swept around the Internet in the fall of 2004 with President George W. Bush and John Kerry singing to the tune of Woody Guthrie’s ‘This Land is Your Land’, the music publisher that owned the copyright to Guthrie’s tune sought to stop distribution of the film. The concern was not that the spoof would directly sub-

---

230 See Alan Sawyer, Repetition, Cognitive Responses, and Persuasion, in Cognitive Responses in Persuasion 237–44 (Richard E. Petty et al. eds., 1981); see also Bargh, supra note 229, at 14 (positing that frequency of exposure to a subject produces positive feelings of trust and liking due to ease of recall that is misattributed to the qualities of the subject itself).


232 See Bill Britt, Disney’s Global Goals, Marketing, May 17, 1990, at 26 (“To avoid overkill, Disney manages its character portfolio with care. It has hundreds of characters on its books, many of them just waiting to be called out of retirement . . . . Disney practices good husbandry of its characters and extends the life of its brands by not over-exposing them . . . . They avoid debasing the currency.”).

233 See supra notes 80–81 and accompanying text; cf. J.C. Herz, Joystick Nation: How Videogames Ate Our Quarters, Won Our Hearts, and Rewired Our Minds 134 (1997) (“Nintendo’s in-house characters are its crown jewels. And the company is fastidious, to the point of paranoia, about safe-guarding their reputations.”).

stitute for sales of albums or singles featuring the song, nor was it having to forego whatever nominal license fee the cartoonists might have paid. Kathryn Ostien, director of copyright licensing for the publisher, told CNN that the harm was the creation of new associations with the song, presumably those of political carping and dissension stating, “This puts a completely different spin on the song. The damage to the song is huge.”

Other copyright owners express similar concerns about the harm of secondary use on the perceptive value of the original. Mattel has sued artists for using its protected Barbie characters in a song about materialism and in fine art photographs or reproductions that placed the doll in erotic or sexualized positions. In these cases, Mattel claimed that the connection of the doll with such unpleasant associations harmed the value of the toy itself in addition to markets for licensed derivatives. The guardians of Margaret Mitchell’s estate, SunTrust Bank, may have objected to Alice Randall’s sexually explicit reworking of *Gone With the Wind* for similar reasons. Likewise, Barney’s owners sued to enjoin a skit wherein the San Diego Chicken beat up an imposter Barney because children in the audience had been frightened and might be turned off the brand. Lever Brothers succeeded in enjoining a parody commercial for a video game in which its Snuggle fabric softener bear was attacked by tanks in part because of the commercial’s association of the Snuggle bear with violence.

---


236 Mattel, Inc. v. MCA Records, Inc., 296 F.3d 894, 899 (9th Cir. 2002) (suing for trademark dilution).

237 Mattel, Inc. v. Walking Mountain Prods., 353 F.3d 792, 796–97 (9th Cir. 2003); Mattel, Inc. v. Pitt, 229 F. Supp. 2d 315, 318 (S.D.N.Y. 2002).

238 *See Walking Mountain*, 353 F.3d at 805; *MCA Records*, 296 F.3d at 902–03.

239 Suntrust Bank v. Houghton Mifflin Co., 268 F.3d 1257, 1282–83 (11th Cir. 2001) (attributing to SunTrust a “vigilance” in policing *Gone With the Wind*’s public image and speculating that the claim was aimed in part at preserving the book’s reputation).


241 Conopco, Inc. v. 3DO Co., 53 U.S.P.Q.2d 1146, 1147 (S.D.N.Y. 1999) (describing the injunction in a later contempt action); *see also* Memorandum of Law in Support of Plaintiff’s Application for a Temporary Restraining Order, a Preliminary Injunction, and an Order for Expedited Discovery at 21, Conopco, Inc. v. 3DO, 53 U.S.P.Q.2d 1146 (S.D.N.Y. 1999) (No. 99 Civ. 10893). Many of these cases are brought under both trademark and copyright. Courts do not always distinguish between these regimes in issuing injunctions. *See, e.g.*, *Conopco*, 53 U.S.P.Q.2d at 1147 (quoting from original TRO order).
2. Overlap with Trademark Dilution

In this way, the extension of copyright to cognition has prompted owners to seek remedies when the “goodwill” associated with expressive elements is threatened. Such copyright suits function similar to actions for trademark dilution. Indeed, many secondary use cases proceed under both trademark dilution and copyright infringement theories.\(^{242}\) Despite the overlap in function, however, copyright protection is more problematic than trademark when used to protect a work’s signaling function because it suppresses a greater range of expressive activity. Trademark protects the ability of merchants to communicate information to consumers about the source of goods or services.\(^ {243}\) Derived from Congress’ power to regulate commerce, it applies only to commercial uses on like goods.\(^ {244}\) By contrast, copyright governs every reproduction or representation of proprietary works, no matter how small or personal. Trademark thus relates only to the ability to choose goods in the marketplace, whereas copyright governs choice between goods and the buyer’s use of those goods.

Trademark protection also differs from copyright in that it is explicitly limited only to “confusing” or demonstrably harmful uses.\(^ {245}\) Trademark infringement cases examine a list of factors such as similarity of the marks and channels of trade to determine whether a junior mark infringes a senior one.\(^ {246}\) Copyright litigation requires no evidence that audiences be confused or misled by similarities between uses. In limited circumstances, owners of famous trademarks may sue in federal court to enjoin commercial use of similar marks on unrelated goods under a “dilution” theory, but to do so they must provide evidence of actual harm to the mark in the form of blurring or tar-

\(^ {242}\) See, e.g., *Walking Mountain*, 353 F.3d at 797; *Giannoulas*, 179 F.3d at 387; *Dr. Seuss Enters., L.P.* v. *Penguin Books USA, Inc.*, 109 F.3d 1394, 1396, 1403 (9th Cir. 1997); *Conodico*, 53 U.S.P.Q.2d at 1147.

\(^ {243}\) See 15 U.S.C. § 1127 (2000) (“The term ‘trademark’ includes any word, name, symbol, or device, or any combination thereof—(1) used by a person, or (2) which a person has a bona fide intention to use in commerce . . . to identify and distinguish his or her goods, including a unique product, from those manufactured or sold by others and to indicate the source of the goods, even if that source is unknown.”).

\(^ {244}\) Id. § 1114.

\(^ {245}\) Id. (describing trademark infringement as uses of marks that are likely to cause confusion among consumers); *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418, 433 (2003) (holding that action for trademark dilution requires proof of actual harm to economic value of a mark).

\(^ {246}\) See *Savin Corp. v. Savin Group*, 391 F.3d 439, 456 (2d Cir. 2004).
nishment of its meaning among relevant consumers. In copyright secondary use cases, courts presume harm to the work through unauthorized use. They place the burden on the user to prove the value of the secondary use through affirmative defenses such as parody. For these reasons, the interpretation of copyright to cover consumer associations provides a kind of super-trademark protection that omits trademark’s safeguards against overreaching.

No court has yet taken the trouble to examine forthrightly whether or not dilution of a work’s meaning properly falls under copyright. Instead, courts have been alternatively receptive or dismissive of the theory depending on their opinion of the value of the secondary use. Where courts find the secondary use to be legitimate criticism of the original or otherwise sympathetic to the message of the new work, they will dismiss the plaintiff’s suit as an attempt to shield the work from effective critique. Although these judges may be right that the public interest in satire outweighs potential harm to a brand, they are wrong when they state that plaintiffs seek only to avoid criticism. The plaintiffs’ concerns are much broader and more banal; they seek to avoid any use that conflicts with a managed brand personality. Indeed, where courts are less sympathetic to the parody at issue or do not find a direct “critique” present, they are more receptive to the idea that loss of control over public presentation works an irreparable harm to the value of the original text. For example, in a

247 Moseley, 537 U.S. at 433. At least one lower court has found this requirement to be looser in the subset of cases where the defendant’s mark is identical or virtually identical to the plaintiff’s. See Savin Corp., 391 F.3d at 452. State standards may also be less stringent, see id. at 456, but not all states have their own dilution acts.

248 See, e.g., MyWebGrocer, LLC v. Hometown Info., Inc., 375 F.3d 190, 192 (2d Cir. 2004); Elvis Presley Enters., Inc. v. Passport Video, 349 F.3d 622, 627 (9th Cir. 2003).

249 See, e.g., Campbell, 510 U.S. at 592–94.

250 Walking Mountain, 353 F.3d at 806; Suntrust Bank, 268 F.3d at 1282–83 (“Suntrust may be vigilant of Gone With the Wind’s public image—but it may not use copyright to shield Gone With the Wind from unwelcome comment . . . .”); Consumers Union of U.S., Inc. v. Gen. Signal Corp., 724 F.2d 1044, 1050 (2d Cir. 1983) (“The fourth factor is aimed at the copier who attempts to usurp the demand for the original work. The copyright laws are intended to prevent copiers from taking the owner’s intellectual property and are not aimed at recompensing damages which may flow indirectly from copying.”) (internal citations omitted).

251 See, e.g., Dr. Seuss Enters., L.P., 109 F.3d at 1403 (considering harm to the substantial “goodwill and reputation” associated with Dr. Seuss in enjoining unauthorized use of Dr. Seuss’s copyrighted Cat in the Hat character in a book about the O.J. Simpson murder trial); Conopco, Inc., 55 U.S.P.Q.2d at 1149 (noting that use of a bear character in enjoined commercial “potentially broadly damaged SNUGGLE ® Bear’s reputation”); United Feature Syndicate Inc., v. Koons, 817 F. Supp. 370, 382 (S.D.N.Y. 1993) (considering effect that large numbers of unauthorized sculptures of cartoon characters would have on public’s
case concerning the use of a James Bond-like character in an advertisement for Honda automobiles, a court enjoined the commercial because of the risk that association of Bond with a “low-end” car like Honda would harm the brand’s upscale licensing strategy.\footnote{Metro-Goldwyn-Mayer, Inc. v. Am. Honda Motor Co., Inc., 900 F. Supp. 1287, 1300–01 (C.D. Cal. 1995).}

For potential users, the uncertainty surrounding this question chills expressive activity and innovation. If secondary authors can convince a court that theirs is a true comment on the original work, the court will discount any harm to brand perception. If, on the other hand, a court rejects the parodic message, loss of control over public associations with the brand most likely will be considered as a market harm. The Supreme Court suggested in a footnote in the \textit{Campbell} decision that, where market harm is unlikely, a secondary use could be less targeted at the original and still qualify as fair use.\footnote{\textit{Campbell}, 510 U.S. at 580 n.14.} This suggestion of flexibility is illusory because from a branding standpoint, all uses of a protected work trespass on potential markets.\footnote{Indeed, in the ten years since the decision, no court has been willing to rest a fair use determination on this basis alone.}

The legal framework thus embodies a one-way assumption for the meaning of expressive texts. According to the law, except in very limited circumstances, owners get to decide how consumers communicate publicly about popular works. This account, based on observations about real property, ignores the participatory nature of informational goods. Such goods derive much of their value from their use and significance to others.\footnote{Pessach, supra note 78, at 1084.} If we allow owners to profit from the positive effects when the public embraces their work, we probably should ask them to bear more risk if audience members grow disenchanted, bored, or rebellious.

One answer to the branding question would be to rule that consumer perception and opinion is completely outside the bounds of copyright. Multiple imitations of a work may “devalue” that work in desire for licensed products); Dr. Pepper Co. v. Sambo’s Rests., Inc., 517 F. Supp. 1202, 1208–09 (N.D. Tex. 1981) (enjoining local television commercial parody of plaintiff’s successful “Be A Pepper” campaign because distractions from the uniqueness and originality of the commercials would logically shorten the life of the ad campaign and damage business goodwill); \textit{cf.} Merges, \textit{Of Property Rules, supra} note 40, at 2659 n.15 (noting copyright’s policy of favoring authors’ reputational interest by ensuring that authors can control all manifestations of a work).
the popular mind, but the work itself is not altered.\textsuperscript{256} The same effect could be found if a purchaser plays a new album over and over to the point where she never wants to hear it again, but the law does not allow the copyright owner to forbid such use.

The problem with discounting consumer perception altogether is the difficulty of separating such a claim from a legitimate cause of action within copyright. The Copyright Act disallows copying of protected elements in any form, absent a few specific exceptions.\textsuperscript{257} Such clear and certain boundaries are thought to facilitate trade and investment in creative works.\textsuperscript{258} Accordingly, an owner need not demonstrate any harm before she is entitled to enjoin the new use or receive damages.\textsuperscript{259} To qualify as a fair use of the work, the burden rests on the secondary user to prove that the use will not harm the value of the original.\textsuperscript{260} Diminishment of consumer demand is relevant to this inquiry.

One could distinguish between harm through direct market substitution and harm through indirect diminishment of brand appeal and audience demand. Indeed, courts often claim to be doing just this.\textsuperscript{261} In a world where every possible derivative belongs to the owner, however, it is difficult to imagine a work that does not directly substitute for

\textsuperscript{256} For an articulation of this specific critique, see Arlen W. Langvardt & Kyle T. Langvardt, \textit{Unwise or Unconstitutional?: The Copyright Term Extension Act, the Eldred Decision, and the Freezing of the Public Domain for Private Benefit}, 5 Minn. Intell. Prop. Rev. 193, 244 (2004). Arlen and Kyle Langvardt find attention to consumer opinion generally inconsistent with copyright’s mandate to promote progress in the arts and sciences. See \textit{id}. They do not examine the specific consequences of including alteration of consumer perceptions as part of an analysis of market harm caused by secondary use.


\textsuperscript{258} See \textit{supra} notes 35–47 and accompanying text.

\textsuperscript{259} See \textit{MyWebGrocer}, 375 F.3d at 192; \textit{Elvis Presley Enters.}, 349 F.3d at 627.

\textsuperscript{260} \textit{Campbell}, 510 U.S. at 590, 594. Courts determine whether an unauthorized use is “fair” upon consideration of four factors. These include: (1) the purpose and character of the use; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion taken; and (4) the effect of the use upon the potential market for or value of the copyrighted work. 17 U.S.C. § 107.

\textsuperscript{261} See, e.g., \textit{Suntrust Bank}, 268 F.3d at 1280–81 (distinguishing between works that “capitalize on . . . notoriety” and those that directly substitute and stating that critical screeds are unlikely to act as substitutes). In between frontal critical attacks and direct substitutes, however, there lies a whole universe of subtle materials and messages that could impair audience demand in a related market. Compare \textit{id} at 1281 n.4 (declining to consider harm of widespread unauthorized sequels on markets for authorized derivatives of a famous book where unauthorized book was, in the court’s opinion, clearly and directly critical of original), with \textit{United Features Inc.}, 817 F. Supp. at 382 (considering harm that widespread unauthorized derivative sculptures could pose to underlying comic book character’s licensing value in all potential three-dimensional or merchandising markets in finding that satirical but not overtly critical fine art reproduction was not a fair use).
some possible licensed use.\textsuperscript{262} Furthermore, the difference between direct and indirect substitution may be illusory for the owner. The owner is just as injured if demand is diverted to a pirate as if demand dries up due to audience fatigue. In reality, the two may boil down to the same thing. The traditional description of a use that substitutes is one that will usurp the demand for the original.\textsuperscript{263} Exhausting a work’s novelty value in service of another product arguably qualifies as “substitute” under this definition.\textsuperscript{264} Excluding consideration of consumer perception then would require at a minimum a significant reconfiguration of the derivative right and the fair use analysis.

This doctrinal confusion reflects the reality that unauthorized dilution poses problems for some works under copyright that do not exist for non-expressive works such as those covered by patent. Take a proven pharmaceutical product like a heart or cholesterol drug. If a pirate attempts to pass off his goods as the proven item, if a disaffected consumer uses the trademark in a disparaging way, or even if a celebrity actor publicly lambasts the drug on a popular morning television show, the drug will work the same way when administered to a patient. The efficacy of the product itself has never been altered. By contrast, the entire purpose of a creative work is to deliver a particular message to an audience. Unfettered secondary use of the creative elements of the work may alter the ability of the original to deliver its intended message. Diminishment of the work’s communicative impact may lessen incentives to invest in the production and dissemination of creative works. In this way, copyright and trademark may legitimately overlap in protecting a work’s expressive value.\textsuperscript{265}

\textsuperscript{262} See Campbell, 510 U.S. at 593 n.24 (noting that because all parodies are also potentially substitutes for a licensed work in a derivative market, it may be difficult to separate permissible harm caused by criticism from impermissible harm caused by substitution). Justice Souter does not clarify whether harm to the work’s retail value through inconsistent use or overuse is properly understood as “effectiveness of critical commentary” or “substitution.” See id.

\textsuperscript{263} See Fisher v. Dees, 794 F.2d 432, 438 (9th Cir. 1986).

\textsuperscript{264} See Patry & Posner, supra note 32, at 1644–45 (suggesting that use that impairs a work’s market or value though “free-riding” is not a fair use).

\textsuperscript{265} Professor Lemley has argued that diminishment of value to producers caused by new entrants to a market is the hallmark of what economists refer to as a “perfect market” and so is not a cause for concern. Lemley, \textit{Ex Ante Versus Ex Post}, supra note 11, at 144. If we equate “value” with price, then Lemley is right. If, however, “value” is taken to mean efficacy or utility of the good, then everyone loses through overuse. Lemley’s account is thus a better fit for goods such as software products than for communicative goods such as texts and music.
Trademark protection by itself may be sufficient to address these harms. Trademark and unfair competition theories are designed to police against secondary uses that deceive consumers or inordinately dilute the expressive value of symbols.\textsuperscript{266} Because trademark reaches only commercial uses, using trademark alone to safeguard the signaling function of a work could by itself provide more latitude for secondary use and recoding.\textsuperscript{267}

This solution is doctrinally satisfying but practically elusive. For one thing, the Supreme Court’s recent decision in \textit{Dastar Corp. v. Twentieth Century Fox Film Corp.} limited the use of trademark theories to protect expressive elements of a copyrighted work.\textsuperscript{268} Furthermore, given the collective clout of the copyright industries, it is highly unlikely that Congress would actually repeal portions of the derivative and distribution rights and the fair use exception necessary to remove copyright from the balance.\textsuperscript{269} Additionally, trademark dilution itself is a notoriously slippery concept and so may not provide much ex ante certainty to encourage secondary users.\textsuperscript{270}

Whatever doctrinal scheme is used to address branding concerns, a primary concern should be to provide secondary authors with clear and predictable guidelines about what kinds of uses are “fair” or exempted. Boundaries that are simple and easy to understand will reduce enforcement and transaction costs for owners, and will provide greater certainty and latitude to secondary users, especially those who


lack the resources and expertise to engage in fine analysis of intellectual property codes. This Article posits that the easiest way to arrive at such certainty, while still balancing concerns of incentive and access, is to take seriously the dynamic, communicative properties of the resource at issue and to tailor secondary use rights accordingly.

C. Consumer Psychology and Consumption

Consumers have interests on both sides of the secondary use debate. Consumers choose between expressive goods at least in part based on brand identity messages or cues from the work’s owner.271 Stable brand and message identities allow for easy and convenient product selection. The process of consuming the work, however, stimulates subjective association, fantasy, and imagery.272 These associations form part of the significance of the work to the user.273 Shared communication of the experience of consumption helps to establish a social meaning of the work that may be different from the meaning put forth by the work’s copyright owner.274 These associations, although stimulated by and inextricably linked to the work, should not be considered part of the original copyright owner’s domain.

1. Choice

Consumer psychology research points to two principal aspects that inform choice of entertainment, or “hedonic,” materials: attention and identity. Choice of any consumer product depends on audience attention and understanding, but these elements have the greatest impact


272 Miller, supra note 271, at 190–91 (discussing from an anthropological perspective the process of transforming through consumption an alienable good to a cultural artifact endowed with particular inseparable connotations); Elizabeth C. Hirschman & Morris B. Holbrook, Hedonic Consumption: Emerging Concepts, Methods and Propositions, 46 J. Marketing 92, 92 (1982).

273 See Hirschman & Holbrook, supra note 272, at 93–95.

274 Cf. Jerome Bruner, Acts of Meaning 12–13 (1990) (stating that culture is shared and participatory; we depend on shared modes of discourse to negotiate the meaning and interpretation of cultural objects); Katherine Nelson, Making Sense: The Acquisition of Shared Meaning 249–51 (1985) (explaining that development of meaning is an interactive process that depends on internal interpretive systems, the context of the use, and how the communication is interpreted within a given community).
for hedonic goods because, unlike food or household staples, they have no independent functions that would require purchase.

a. **Attention**

The most salient feature of the current information economy is the sheer number of works competing for the attention of an audience. Twenty years ago, producers of expressive goods faced a marketplace consisting of relatively “few outlets and essentially a captive audience.” A scarcity of informational product existed in the face of mass demand. New methods of information distribution, such as television, the Internet, and portable digital devices, have changed all this. Information can reach us at our desks, at home, and as we travel between the two. We are also subject to an expanding volume of media products. This wealth of information creates a poverty of attention.

A familiar text, work, or personality allows consumers to pick out products of interest and tune out the rest. Buying on the basis of gut feelings about a product, such as a general positive attitude or “liking” created by brand identity, saves consumers cognitive energy. They can sidestep feelings of doubt and hesitancy that naturally result from paying closer attention to any consumptive choice.

The importance of attention helps to explain the tenacity of parody even in the tightly controlled proprietary marketplace. By placing an older work in a new context, either in an advertisement or in a product itself, producers offer familiarity and surprise, two elements likely to captivate an audience. In a consumer research study exam-
ining advertising effectiveness, researchers found that people spend more time on and have greater memory for advertising messages that use familiar elements in unfamiliar, unexpected ways. Studies show that humor also can increase a consumer’s response to a message. Humorous messages command more attention, but they also act “as a distraction, disarming possible criticism.”

One way to view parody then is as a device that fools consumers into paying attention to something they would ordinarily disregard.

This impact persists whether or not the secondary user is parodying the work, using it to satirize something else or just trying generally to get attention. The Supreme Court’s distinction between productive uses and uses just to get attention in this respect creates a false dichotomy. Every secondary user, parodist or not, will attract a greater audience for a message by tying it to a well-known work than they would by distributing the message alone. The problem is that everyone cannot use expressive works at once. Symbols in the form of characters, sounds, and images will cease to catch the attention of consumers if they see or hear them everywhere. Consumers have an interest in controlled use of branded properties so that they can allocate their attention efficiently.

2. Identity

After awareness, consumer research suggests that the choice to consume a given work depends on that work having some salient meaning or symbolic value. This is what Posner and Landes mean when they argue that the transformation of Mickey Mouse from Dis-

---

284 Pieters, et al., supra note 283, at 773; see also Thomas H. Davenport & John C. Beck, The Attention Economy: Understanding the New Currency of Business 195 exh.11–1 (2001) (finding that factors that attracted attention to email messages included that the message concerned topics of personal interest and that the content was presented in a new, unusual or unique way).

285 See Lee & Mason, supra note 283, at 168.

286 O’Shaughnessy & O’Shaughnessy, supra note 210, at 132; Lee & Mason, supra note 283, at 168.

287 See Janos Laszlo, Cognition and Representation in Literature: The Psychology of Literary Narratives 96 (1999); Richard P. Bagozzi et al., The Role of Emotions in Goal-Directed Behavior, in The Why of Consumption: Contemporary Perspectives on Consumer Motives, Goals, and Desires 37, 50 (S. Ratneshwar et al. eds., 2000) (hypothesizing that many consumer behaviors, such as seeing movies, are the result of anticipated emotions); Morris B. Holbrook & Elizabeth C. Hirschman, The Experiential Aspects of Consumption: Consumer Fantasies, Feelings, and Fun, 9 J. Consumer Res. 132, 138 (1982).
ney’s grinning spokescharacter to a raffish derelict or hen-pecked husband will harm the value of the character himself.\textsuperscript{288} Popular characters, sounds and images provide useful information. We can expend less effort choosing between films for example, if one is a sequel or derivative. We do not need to spend time and energy researching the plot of the film; we know we are interested in the character or story.

Specifically, consumer research shows that we choose hedonic goods based on the experience we think they will offer us.\textsuperscript{289} Attractive qualities in expressive works include works that we think will be transporting, that contain a character or storyline in which we would like to envision ourselves, or that offer access to a desired group or situation.\textsuperscript{290} In this way, most consumers interact with expressive products not with the aim of transforming the product, but with the desire to be transformed by the product.\textsuperscript{291} For example, we tend to choose to purchase music based on its ability to draw us in and offer an absorbing experience.\textsuperscript{292}

Once again, secondary users benefit by defining their own work in opposition to a previous one. Communicating a salient identity to distracted consumers is difficult. Producers seeking to market to one target segment can use existing goods as informational shortcuts. For example, the popular Middle Eastern brand Mecca Cola is beloved because it is marketed explicitly as an opportunity to reject American imperialism in the form of the American brand standard Coca-

\footnotesize
\textsuperscript{288} See Landes & Posner, supra note 7, at 487–88.
\textsuperscript{289} See Laszlo, supra note 287, at 96 (1999); Bagozzi et al., supra note 287, at 50 (hypothesizing that many consumer behaviors such as seeing movies are the result of anticipated emotions); Holbrook & Hirschman, supra note 287, at 138.
\textsuperscript{291} Hughes, supra note 7, at 957.
Cola. The video game company that based a series of ads on blowing up Lever’s Snuggle fabric softener bear gained instant credibility with adolescent male gamers. These products catch the eye of a target market by repurposing symbols of lifestyles, the cuddly safety of childhood, or the brash consumerism of the United States, about which their audience is ambivalent.

The same effects hold true when previous works are incorporated into the work itself. The rap group 2 Live Crew identified itself as dangerous, outside, and urban by reworking a nostalgic seventies tune. Alice Randall sent a message about the prominence of black Americans in culture by upending the racial hierarchies in the original Gone With the Wind. Each artist made their message more interesting by framing it in terms of a text with which a significant part of the audience was already invested.

Overuse or inconsistent use of such signifiers, however, ruins their informative value and impedes their utility for audience choice. The works themselves may be less desirable for consumption if their meanings are altered. For example, music that has been used for a variety of commercial purposes may cease to transport listeners easily. Characters that act every which way will make poor vehicles for individual fantasy. Overuse and widespread conflicting uses may interfere with the ability of users effectively to choose works to consume. Although owners may have no right to dictate the meaning of their works to members of the audience, they do have some claim to preserving the ability of audiences to choose to engage with their works.

Others have argued that widespread use will have mostly positive effects on audience desire and the value of individual works. For example, the value to me of seeing a certain film, The Aviator, say, increases as more of my friends see it. Seeing the film then provides not only temporary enjoyment, but a basis for connection and discussion with a desired peer group. This is also true of scholarly works. The value of this Article increases the more people who read it and

293 O’Shaughnessy & O’Shaughnessy, supra note 210, at 88.
295 See generally Campbell, 510 U.S. 569.
296 See generally Suntrust, 268 F.3d 1257.
297 See Hughes, supra note 7, at 942–43.
299 See id.
quote its ideas. Such positive externalities are termed “network effects.” While this is a good argument for free dissemination of exact copies of works, it is more complicated when it comes to derivatives. The theory depends on my friends and me all having seen the same version of _The Aviator_ or we will have no basis for comparison. Network effects fail if we cannot easily exchange information about our shared experience.

To be sure, some derivatives do increase the value of originals. Sales of books tend to increase after the release of film and DVD versions, for example. In such cases, however, the derivative is complementary to the original. It is unclear what the release of badly made, offensive, or critical derivatives might have on consumer perceptions of and desire for an original work. Furthermore, copyright and other proprietary schemes ensure that only one authorized derivative is released in each category. Without copyright, audiences could be bombarded with multiple sequels to popular films, film versions of popular books, and commercials featuring characters from each. The ubiquitous presence of multiple texts may decrease the desirability of every part of the franchise even though individual uses may have found an audience in isolation. Consumers will also have to spend more time and energy sorting through competing messages, something they may be unwilling to do when it comes to works chosen for escapist, experiential reasons.

3. The Audience’s Personal Use of Expressive Works

Once a consumer has access to a work or has it in his possession, the relationship changes. Cultural works provide a variety of complex and important functions for users. These experiences extend beyond simple entertainment or desire to pass time. Users engage with cultural works as a way of projecting and exploring their idealized “true” selves. Engagement with adventure and risks not present in daily

---

300 See Pessach, _supra_ note 78, at 1085 (describing cultural network effects).
301 Sales of books from the Lord of the Rings trilogy shot up after the release of film and DVD versions, for example. Amazon Sales Rank for Lord of the Rings, http://www.amazon.com (search “Books” for “Lord of the Rings”; sort by “Bestselling”; select “J.R.R. Tolkien Boxed Set (The Hobbit and The Lord of the Rings) [BOX SET]”; then scroll down to “Product Details”).
302 See Abramowicz, _supra_ note 125, at 5, 23.
303 See Holbrook & Hirschman, _supra_ note 287, at 138; see also Russell et al., _supra_ note 215, at 275, 276 (listing reasons why people watch soap operas).
304 See Eric J. Arnould & Linda L. Price, _Authenticating Acts and Authoritative Performances: Questing for Self and Community, in The Why of Consumption: Contemporary Per-
life allows for emotional release and fantasy. Such works also can assist in acclimating to new situations or establishing cultural norms. Watching a film about a painful situation, such as the divorce in Kramer v. Kramer, can help individuals process difficult experiences and painful emotions, for example. The knowledge that many others have experienced the same works can inform our sense of appropriate reactions to novel situations. Even works of fashion dictate rules of appearance that consumers can choose whether and how to adopt. Cultural works also assist in defining and establishing social networks. Common consumption of the new hit show or novel creates a shared experience.

spectives on Consumer Motives, Goals, and Desires, supra note 287, at 140, 145; Russell et al., supra note 215, at 277; cf. Bruner, supra note 274, at 33, 101 (stating that the self is shaped in reference to cultural systems of interpretation).

Hirschman & Holbrook, supra note 272, at 96; see also Arnould & Price, supra note 304, at 150 (reporting that consumers of rafting trips gain self-assurance and confidence through memories of the experience); Melanie C. Green et al., The Power of Fiction: Determinants and Boundaries, in The Psychology of Entertainment Media: Blurring the Lines Between Entertainment and Persuasion, supra note 215, at 161, 169 (describing how participants who felt transported into a story came to see characters as friends and remembered narrated events as real).

See Maria Kniazeva, Between the Ads: Effects of Nonadvertising TV Messages on Consumption Behavior, in The Psychology of Entertainment Media: Blurring the Lines Between Entertainment and Persuasion, supra note 215, at 213, 218, 220 (noting that media sorts reality into meaningful social categories that consumers can use to interpret their daily lives and can help to acculturate foreigners to new environments); cf. Bruner, supra note 274, at 68 (stating that human propensity to share stories and find interpretations within dominant moral and institutional schemes is one of the most powerful forms of social stability).

Hirschman & Holbrook, supra note 272, at 96.


See Miller, supra note 271, at 209–10, 212; Hirschman & Holbrook, supra note 272, at 99; Russell et al., supra note 215, at 279; cf. Arnould & Price, supra note 304, at 154–58 (arguing that certain acts of consumption, such as viewing or participating in Christmas pageants, river-rafting trips, tourism, and football games, contribute to self-conception and connection to community).

communities or shared rituals, expressive goods provide outlets for interacting with others.\textsuperscript{312} At the same time, the act of consuming a work alters its meaning for the individual. When reading, watching a performance, or listening to music, users not only process the sensory attributes of the work itself, but also generate internal, multisensory imagery.\textsuperscript{313} This can include real memories triggered by association, fantastical imagery inserting ourselves in or in relation to the work, and pure affective arousal.\textsuperscript{314} In this way, consumers of a work encode their experience of the work differently in memory.\textsuperscript{315} Users enhance their experience of a work by sharing and contrasting their own consumptive experience with others.\textsuperscript{316} Some claim that this ability to communicate and share one’s experience with the work is a crucial step in processing and internalizing memory of the work.\textsuperscript{317} At one extreme, people strongly affected by experiences with expressive works may join fan clubs, participate in fantasy reenactments, or create “fan fiction” in the form of unauthorized derivatives.\textsuperscript{318} Only a copyright zealot would suggest that an individual’s experience of a text somehow belongs to the copyright owner. And yet, many of the lawsuits surrounding secondary use of expressive works


\textsuperscript{313} Laszlo, supra note 287, at 98; Hirschman & Holbrook, supra note 272, at 92–94.

\textsuperscript{314} See Laszlo, supra note 287, at 96–98; Athinodoros Chronis & Ronald D. Hampton, Living in Another World: The Role of Narrative Imagination in the Production of Fantasy Enclaves, 31 ADVANCES IN CONSUMER RES. 193, 193 (2004); Green et al., supra note 305, at 169.

\textsuperscript{315} See, e.g., Walter Kintsch, The Representation of Meaning in Memory 107–08, 115–16 (1974) (finding that representations are stored in memory on a semantic rather than literal basis and individual memory of texts may differ based on differences in environment and presentation). Cognitive research posits that people encode stories in the form of schemas or scripts that assimilate the details of the story with preexisting knowledge about its subject. Robinson-Riegler & Robinson-Riegler, supra note 29, at 464–65. People’s operative schemas, however, may differ depending on their place in the culture. See Bruner, supra note 274, at 64; Laszlo, supra note 287, at 103–07.

\textsuperscript{316} See Tushnet, supra note 100, at 545–46; see also Bruner, supra note 274, at 13, 33, 68 (arguing that as cultural beings our way of life depends on shared meaning and concepts which are negotiated through discourse about interpreting narratives and symbols); Wolf, supra note 206, at 38 (entertainment products put people on the same wavelength and “replace a sense of shared community”); Brock & Livingston, supra note 311, at 272 (concurring with Wolf).

\textsuperscript{317} See Gillespie, supra note 29, at 148 (arguing that memory and cognition are themselves participatory and that through imaginative reconstruction and critical reflection with others we come to understand ourselves and our relationships); see also Bruner, supra note 274, at 33.

\textsuperscript{318} Russell et al., supra note 215, at 279; Tushnet, supra note 103, at 651–58.
result from a user’s attempt to communicate a subjective understand-
ing of the work to others. Alice Randall claimed that she wrote her “parody” of *Gone With the Wind* as a way of “dealing with my response to a text, portraying my response to a text.” Her response was also intended to communicate what she felt was a common view among African-Americans explaining, “I think...the time has come for America to understand how an African-American woman, and many African-Americans, view the book that has influenced our country’s culture and how we view ourselves as a country.” Fans who “dress up” as comic book characters online or who publish unauthorized sequels or spin-offs are acting similarly. These individuals are making use of what is in some sense their own experience. It can be argued that the ability to enjoy the consumptive experience and share it with others is what consumers believed they paid for when they bought access to the work.

What is needed then is a theory that can preserve the audience’s ability to fully use expressive works while protecting at least some of the identity that allows consumers to distinguish between works.

### III. Toward a Balanced Test for Secondary Use

Contrary to the overgrazing paradigm in which all secondary uses will equally undermine audience conceptions, cognitive research suggests that all secondary uses do not impact audience members equally. Although the dynamics of memory and perception are complicated, four well-established doctrines may aid in drawing lines between intangible entitlements and audience reuse rights. These are resistance, frequency effects, source effects, and hierarchy of processing.

First, our general attitudes towards iconic works, which are most often the subject of disputes, are resistant to change. We are more likely to pay attention to works that conform to what we already believe and to discount works that conflict with our attitudes. This

---


320 *Id.*

321 For example, the designer of the City of Heroes game objected to by Marvel Comics sought to “make the experience of playing City of Heroes just like my childhood comic-book experience” by allowing users to design their own fantasy superheroes based on actual comic book characters. Michael Lafferty, *Cryptic’s Jack Emmert Takes a Few Moments from Polishing Issue 3 of City of Heroes to Talk About What Lies Ahead for Gamers*, PC GAMEZONE, http://pc.gamezone.com/news/12_10_04_06_16PM.htm (last visited Sept. 11, 2005); *see also* Tushnet, *supra* note 103, at 657.

322 *See* Schumann, *supra* note 277, at 235.
suggests that people attracted to “parodic” reworkings of texts already share the attitude of the new version, while those who prefer the old will disregard it.

Second, as we cannot possibly pay attention to all of the messages, content, and persuasion hurtling our way from multiple sources, we use processing cues, or heuristics, to determine which messages deserve attention and concentration. Several common processing heuristics, such as source effects, frequency of exposure, and level of processing effort, may help us to better delineate between secondary uses that are likely to distort the identity of works for audiences and those that will not. These processes can thus inform the boundaries we draw around rights to expressive works.

1. Resistance

The first doctrine of relevance to secondary use is attitude resistance. Consumer purchasing decisions are to some extent based on attitudes.323 We buy what we have a positive feeling toward and we avoid what we dislike.324 Once we have formed strong attitudes about an object, we tend to resist change.325 This may be because often certain attitudes are linked as part of an overall self-schema.326 People who live in big cities, for example, tend to prefer different cars, clothes, political candidates, and cultural works than those from rural areas.327 Changing one’s belief may call into question the entire self-schema.328 To avoid dissonance, we tend to ignore or discount messages that call our attitudes into question and we give greater attention to those that reinforce our beliefs.329 The more familiar and rote a perception or belief is, the less likely we are to alter it either consciously or unconsciously through new information.330 While this state

325 Eagly & Chaiken, supra note 323, at 559; see Green et al., supra note 305, at 172–73.
326 See Schumann, supra note 277, at 234–35.
328 Eagly & Chaiken, supra note 323, at 584–89; see Schumann, supra note 277, at 235.
329 Eagly & Chaiken, supra note 323, at 590–608.
of affairs may undercut the possibility of cultural dialogue through reworkings of popular texts, it also suggests that as a practical matter such reworkings may not have much affect on overall audience perception. People who have a long history of positive relations with a work, such as an iconic novel like *Gone With the Wind*, are likely to discount any information that might persuade them to change their attitude. Those already ambivalent about the work and its portrayal of race relations will enjoy seeing these problems exposed through a work like Randall’s *Wind Done Gone*. Similarly, people who love Dick and Jane for their innocence will tend to ignore any negative connotations from the suburban Yiddish version. Those who have always found Dick and Jane a little saccharine and stiff will enjoy seeing the duo confront a more dangerous world than they have been used to. The newer works do not so much steal an audience from the original as they allow a richer discussion for those so inclined. These works are unlikely to chip away much at underlying attitudes about iconic originals because those attitudes are so rote and well-rehearsed. Furthermore, continued marketing and promotion of the original versions will only reinforce the strength of dominant consumer attitudes.

With newer works or ones that are not yet familiar, memory and attitude is more vulnerable to change. Information to which we are frequently exposed and of which we are often reminded is less susceptible to distortion. Newly formed perceptions, by contrast, are more vulnerable to alteration by conflicting information. These findings support suggestions that newer works deserve greater protection than more established ones. Perhaps owners deserve an absolute protection from unlicensed commercial derivatives for one year following publication, for example, with wider use permitted later.

2. Source Effects

Research on perception and memory also reveals that users process messages differently depending on the source. With so many mes-

---

331 *Cf.* Schumann, *supra* note 277, at 235–36 (using studies in the sociology of media consumption to argue that people restrict their exposure to dissonant media messages).


333 *Id.*

334 *Id.* at 256, 265.

sages clamoring for our attention, we decide which to pay attention to partially based on the credibility or attractiveness of the source. A source’s credibility refers both to its perceived expertise on a topic and its trustworthiness as a communicator. Where consumers have little attention or involvement with the subject of a message, as they might with advertising or promotion of a new media product, they tend to evaluate the message based on peripheral cues such as the message source. They more readily accept information from a source perceived as credible and discount information from sources perceived as biased or untrustworthy. For example, one experiment comparing high to moderate credibility sources asked individuals to rank several abstract modern poems based on use of alliteration. They were then shown an article that identified the poem they ranked second to last as a superior example of alliteration and were then asked to rank the poems again. Those who were told that the article had been written by

336 See Davenport & Beck, supra note 284, at 195; Shelly Chaiken, Heuristic Versus Systemic Information Processing and the Use of Source Versus Message Cues in Persuasion, 39 J. Personality & Social Psychol. 752, 753, 763 (1980) (finding that perceptions of a source’s expertise affected opinion of a message and hypothesizing that source cues affect opinion change because they influence both attention to a message and predisposition to its content).


338 See Shelly Chaiken & Durairaj Maheswaran, Heuristic Processing Can Bias Systemic Processing: Effects of Source Credibility, Argument Ambiguity, and Task Importance on Attitude Judgment, 66 J. Personality & Social Psychol. 460, 466, 469 (1994) (describing experiment in which students evaluated an unfamiliar product based on descriptions from sources of varying credibility); see also Richard E. Petty & John T. Cacioppo, Communication and Persuasion: Central and Peripheral Routes to Attitude Change 54, 142 (1986) (finding that individuals with low involvement in subject matter are more likely to use cognitively less taxing “peripheral processes” such as source expertise to evaluate the validity of a message).

339 Chaiken & Maheswaran, supra note 338, at 464; Hass, supra note 337, at 154–55 (noting, however, an exception where the subject is already predisposed to believe in or trust the opinion expressed); see also Paulette M. Gillig & Anthony G. Greenwald, Is It Time to Lay the Sleeper Effect to Rest?, 29 J. Personality & Social Psychol. 132, 135 (1974) (noting tendency to discount information from low credibility sources); Rochelle Lynn Chaiken, The Use of Source Versus Message Cues in Persuasion: An Information Processing Analysis 17–26 (Dec. 1977) (on file with the author) (summarizing research studies); cf. Tulin Erdem & Joffre Swait, Brand Credibility, Brand Consideration, and Choice, 31 J. Consumer Res. 191, 192, 196–97 (2004) (finding, with some variation across product class and individual expertise, that consumers are more likely to include high-credibility brands than low-credibility brands in initial purchase consideration sets and to choose such brands for purchase because highly credible brands have lower perceived risks and information costs).

340 Hass, supra note 337, at 158.

341 Id.
T.S. Eliot were likely to change their opinion of the poem’s merit.\textsuperscript{342} Those who were told the article was written by a college student seeking a teaching position in English were much less inclined to change their opinion than the former group.\textsuperscript{343}

Source credibility effects can explain the phenomenon observed by Tushnet and others that users seem not to mind unauthorized reworkings of popular texts in the form of fan fiction or parody so long as one “orthodox” version exists.\textsuperscript{344} It may be that consumers are perfectly capable of contextualizing reworkings of expressive texts if they have sufficient information about the source. Those who prefer the authorized version will stick with that; those looking to explore alternatives can do so freely so long as unauthorized versions are clearly marked and do not contain misleading source information.\textsuperscript{345} Thus, works clearly identified as emanating from an unauthorized source, either in a labeled “parody” work such as the Yiddish with Dick and Jane book, or in the context of a sketch comedy show, for example, should be expected to have less involuntary impact on consumer attitudes than confusing or ambiguous secondary uses.

3. Frequency Effects

Frequency and repetition, by contrast, can confuse and exhaust audiences. As noted above, after an initial level of familiarity, overexposure to a communication elicits boredom and dislike.\textsuperscript{346} In a study of attitudes towards television commercials, research demonstrated that greater exposure to similar commercials decreased liking and positive feelings both for the commercials and the products advertised.\textsuperscript{347} Variations in the commercials designed to enhance attention had little ame-

\textsuperscript{342} Id.
\textsuperscript{343} Id.; see also J.P. Das et al., Understanding Versus Suggestion in the Judgment of Literary Passages, 35 J. Abnormal & Social Psychol. 547, 547–561 (1967).
\textsuperscript{344} See Tushnet, \textit{supra} note 103, at 672–73; see also Goldberger, \textit{supra} note 209, at 353 (noting anecdotal evidence that audiences do not mind inconsistencies across works so long as one version is deemed official or canonical).
\textsuperscript{345} By “source” information, I do not mean that in the ordinary case a simple disclaimer of affiliation would suffice. As demonstrated in the trademark context, disclaimers can enhance confusion by associating the defendant’s goods directly with the plaintiff. See Mitchell E. Radin, Disclaimers as a Remedy for Trademark Infringement: Inadequacies and Alternatives, 76 Trademark Rep. 59, 65 (1986). In this context, “source” would have to include at a minimum (a) clear and conspicuous indication that the work is “unauthorized” and (b) the identity of the new author if it is likely to be meaningful to consumers.
\textsuperscript{346} See \textit{supra} notes 231–32 and accompanying text.
\textsuperscript{347} Calder & Sternthal, \textit{supra} note 231, at 185.
liorating effect.\textsuperscript{348} Similar effects were found in another study using non-commercial radio messages.\textsuperscript{349} Repeated exposure to a message may also lead to involuntary changes in perception, even if the recipient initially discredited the source as not credible or biased.\textsuperscript{350} In this way, frequent exposure may override the efficacy of other informational cues such as source and so confuse consumers as to authorized and illicit interpretations.\textsuperscript{351} To avoid this aspect of “overgrazing,” secondary uses most likely to distort audience perception should remain subject to property remedies like an injunction.\textsuperscript{352} These would include mass media advertisements incorporating proprietary works and other mass promotions of secondary works that reach audiences unbidden and have the potential to distort audience memory for a work.\textsuperscript{353} So, for example, secondary use in pop-up ads, television advertising, flyers, posters, and perhaps even pervasive forms of mass media distribution such as general radio promotion for songs or wide release for films would remain subject to injunction. Works which must be sought out to be consumed such as books, plays, website parodies, or those that are subject to small-scale distribution, such as casual emails, could not be

\textsuperscript{348} Id.
\textsuperscript{349} Cacioppo & Petty, supra note 231, at 99.
\textsuperscript{350} See Lynn Hasher et al., Frequency and the Conference of Referential Validity, 16 J. VERBAL LEARNING & VERBAL BEHAV. 107, 107–12 (1977) (noting that experiment subjects were more likely to accept repeated statements about subjects such as politics, science and art as true over similar non-repeated statements); see also Scott A. Hawkins & Stephen J. Hoch, Low-Involvement Learning: Memory Without Evaluation, 19 J. CONSUMER RES. 212, 214 (1992) (stating that consumers accept repeated and familiar messages as true without evaluating the content of the message); Herbert E. Krugman, The Impact of Television Advertising: Learning Without Involvement, PUB. OPINION Q. 349, 354 (1965); cf. Alice H. Eagly & Shelly Chaiken, Attitude Strength, Attitude Structure and Resistance to Change, in ATTITUDE STRENGTH: ANTECEDENTS AND CONSEQUENCES 413, 427 (Richard E. Petty & Jon A. Krosnick eds., 1995) (“[W]ell-documented successes in changing strong attitudes have involved bombarding the targeted people with a large amount of information consistent with the desired attitude(s).”); Sawyer, supra note 230, at 254 (observing that although initial responses to obviously manipulative messages may be negative, repeated exposure to the messages may allow for dissipation of the negative responses and subsequent re-evaluation of the message based on quality and believability of the message).
\textsuperscript{351} See Kathryn A. Braun, Postexperience Advertising Effects on Consumer Memory, 25 J. CONSUMER RES. 319, 332 (1999) (finding that consumers may confuse the true nature of product experience with a subsequent ad-induced description of it).
\textsuperscript{352} Cf. id. at 321, 332–33 (finding that subsequent misinformation can alter beliefs without producing a conscious sense of confusion); Gavan Fitzsimons et al., Non-Conscious Influences on Consumer Choice, 13 MARKETING LETTERS 269, 273 (2002) (same).
\textsuperscript{353} See Campbell v. Acuff-Rose Music, 510 U.S. 569, 580 n.14 (1994) (suggesting that works with “wide dissemination” are more likely to substitute for the original than works with “minimal distribution”). See generally Braun, supra note 351.
enjoined. Owners would thus retain strong property boundaries against the most harmful types of secondary use.

4. Hierarchy of Processing

A fourth theory relevant to this area is research postulating high and low systems of cognitive processing. This research argues that individuals process certain types of information using more or less cognitive effort.\textsuperscript{354} Systemic processing involves relatively high degrees of attention and logical processing of message content.\textsuperscript{355} By contrast, we use low-involvement processing for messages that seem unimportant, trivial or when we lack sufficient attention or expertise to logically evaluate the subject matter.\textsuperscript{356} In low-involvement processing, we use cues such as source or familiarity to determine our attitude toward the communication and we tend to ignore the details of precise information contained in the message.\textsuperscript{357} For this reason, low-involvement messages can actually have greater unconscious impact on our attitudes.\textsuperscript{358} Although we have not consciously evaluated the truth or credibility of such messages, they remain stored in memory and may unconsciously impact our opinions and desires the next time we consider a topic related to that message.\textsuperscript{359} So, for example, when we see an advertisement describing Morton Salt as “easy to pour,” we do not really pay attention, but we store a few bits of information from the ad without any conscious elaboration.\textsuperscript{360} Over time, and with repeated exposure, Morton Salt becomes associated in our minds with “easy to pour.”\textsuperscript{361}

The theory has relevance for secondary use because certain types of expressive works tend to be processed along one path or the other. Depth of processing to some degree depends on individual goals in

\textsuperscript{354} See, e.g., PETTY & CACIOPOPO, supra note 338, at 54; CHAIKEN & MAHESWARAN, supra note 338, at 54; CHAIKEN & MAHESWARAN, supra note 338, at 460.

\textsuperscript{355} CHAIKEN & MAHESWARAN, supra note 338, at 460; CHAIKEN, supra note 336, at 752.

\textsuperscript{356} PETTY & CACIOPOPO, supra note 338, at 54; CHAIKEN & MAHESWARAN, supra note 338, at 460; CHAIKEN, supra note 336, at 752; HAWKINS & HOCH, supra note 350, at 213.

\textsuperscript{357} CHAIKEN & MAHESWARAN, supra note 338, at 460; see PETTY & CACIOPOPO, supra note 338, at 54; HAWKINS & HOCH, supra note 350, at 214.

\textsuperscript{358} HAWKINS & HOCH, supra note 350, at 213–14; KRUGMAN, supra note 350, at 354.

\textsuperscript{359} HAWKINS & HOCH, supra note 350, at 213–14; see also ROBERT HEATH, THE HIDDEN POWER OF ADVERTISING: HOW LOW INVOLVEMENT PROCESSING INFLUENCES THE WAY WE CHOOSE BRANDS 95 (2001) (noting that a marketing campaign that promoted the “refreshment” aspect of Heineken contributed to consumers subconsciously associating Heineken with the concept of refreshment).

\textsuperscript{360} HAWKINS & HOCH, supra note 350, at 214.

\textsuperscript{361} Id.
approaching a type of communication.\textsuperscript{362} Those actively seeking information about an imminent decision are more likely to use systemic processing than those seeking diversion or pleasure.\textsuperscript{363} Certain types of works, however, are susceptible to generalization. Longer literary works, for example, demand greater cognitive involvement regardless of individual goals.\textsuperscript{364} Quick and simple images such as cartoons or catchy melodies, on the other hand, are more likely to receive heuristic processing.\textsuperscript{365} Although knowledge about a source may lead us to discount such works when they emanate from authors other than the original and so blunt their impact, this effect may be incomplete.\textsuperscript{366} If we want to provide maximum protection to copyright owners, then we would want to be more concerned with likely low-involvement processing works than with dramatic or literary works that demand more sustained cognitive attention. In this respect, the older copyright acts, which provided for different bundles of rights depending on the type of work, may have more internal logic than the current system of mostly uniform rights.

\textsuperscript{362} Eagly & Chaiken, supra note 323, at 330.

\textsuperscript{363} See, Richard E. Petty, John T. Cacioppo, and David W. Schumann, Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement, 10 J. Consumer Res. 135, 138 (1983); Richard C. Vincent & Michael D. Basil, College Students’ News Gratifications, Media Use and Current Events Knowledge, 41 J. Broadcasting & Electronic Media 380, 380–89 (1997) (college students approaching graduation consumed all types of media more instrumentally to gain knowledge about the world they were about to enter).

\textsuperscript{364} See, e.g., Eysenck, supra note 29, at 77–78 (tasks that involve processing of meaning receive greater depth of processing); Fergus I.M. Craik & Robert S. Lockhart, Levels of Processing: A Framework for Memory Research, 11 J. Verbal Learning & Verbal Behavior 671, 680 (1972) (hypothesizing that tasks that require analysis of meaning of words lead to greater memory retention due to greater depth of processing required by semantic elaboration as opposed to syntactic or structural judgments).

\textsuperscript{365} Cf. Hawkins & Hoch, supra note 350, at 223 (advertisers prefer jingles and rhymes because they increase memory without stimulating elaborative processing). Ease of consumption is a general motivating factor for low-involvement consumers. See id.

\textsuperscript{366} For example, some studies have found a “sleeper effect” for messages from low credibility sources. Darlene B. Hannah & Brian Sternthal, Detecting and Explaining the Sleeper Effect, 11 J. Consumer Res. 632, 632 (1984). Although such messages are initially discounted, over time listeners may retain the information while forgetting its source, leading them to gradually accept the message over time. Id. Other studies, however, have failed to replicate this effect, or have found it to be very limited. Gillig & Greenwald, supra note 339, at 138–39.
5. Implications for Judicial Analysis of Secondary Use Under Copyright

The virtue of adopting an analysis based on consumer choice is to remove analysis of artistic meaning from the judicial sphere and to provide clearer rules to secondary users and to owners. Judges are not well-equipped by training or experience to arbitrate over the objective meaning of cultural works. They can, however, with relative ease determine whether a work is old or new, whether source information is present, whether a message has been mass-advertised or otherwise pushed on large segments of the audience, or whether a work is a book or a poster. In this way, judges need not learn the underlying science to incorporate insights from cognitive research as well as economics. Such an analysis would also provide secondary users with more certainty by providing a clearer indication of works that cannot be enjoined.

Judges could easily incorporate this analysis within the existing fair use framework. The fourth factor under the current fair use test asks judges to determine the effect of the secondary work on the market for the original.367 No standards currently exist by which to measure market harm. Although many have argued that only uses that substitute for the original in the market should count, widespread licensing of expressive works challenges our definition of what the “market” should be. Is it only uses for which the owner has already licensed the original? Uses that the owner might want to license in the future?368 Or, as many owners argue, any use that lowers the licensing value of the brand more generally? The above rules directly address the issue of harm to expressive value, and allow judges a cleaner framework for adjudicating the fourth factor in secondary use cases.369 Such a test would remove the need for a special “parody” ex-

---

367 *Campbell*, 510 U.S. at 590.
368 See Goldstein, *supra* note 65, at 232–34 for an argument in favor of this broader definition.
369 As an alternative, such factors could be used to frame a more contextual inquiry into “substantial similarity” between an original and a derivative work. Because these factors help distinguish between uses that consumers will find confusing and those that they will contextualize effectively, they are relevant to the question of what kinds of unauthorized derivative works overstep the boundary between permissible conjuring of an original work and impermissible “copying” of that work. See Laura G. Lape, *The Metaphysics of the Law: Bringing Substantial Similarity Down to Earth*, 98 DICK. L. REV. 181, 194–202 (1994) (arguing generally that a finding of substantial similarity should depend on likelihood of harm to the owner’s incentive to create). These factors might also be useful in more narrowly defining trademark uses that potentially “dilute” older marks.
ception except perhaps in the relatively rare cases where artists seek to mass disseminate a commercial parody to an involuntary audience.

Proponents of secondary use might argue that restrictions on methods of promotion or types of work unfairly burden an important type of expression and interfere with its finding an audience. To the extent that restrictions on dissemination and advertisement burden secondary works, this is not necessarily a bad thing. The disadvantage of allowing greater personal and commercial use of popular copyrighted works is that it becomes cheaper to obviously piggyback on existing works than to create more original texts. In an attention economy, works positioned off of popular brands will have an easier time getting attention and establishing a personality relative to consumer expectations. Limits on advertising and methods of distribution may help to remedy this imbalance and ensure sufficient investment in more original works. Compulsory license fees or actions for profit allocation, such as Netanel’s or Rubenfeld’s schemes, may also be appropriate in certain commercial contexts.

Conclusion

Copyright’s cognitive elements have gone unexamined for too long in favor of an analysis grounded in real property theory. Rules based on pastures and cattle have limited force when applied to rights in information. A more explicit grounding of copyright principles in the mechanics of human cognition, similar to the types of analysis commonplace under trademark law, provides a better balance between rights of owners and needs of users. In many ways, such a theory is no more controversial than suggesting that the fee simple estate in real property conform to the known dynamics of physical space. Courts once considered air rights part of a landowner’s domain. This definition changed when air traffic increased the social value of public air space to the point where it outweighed the private need for such rights. Similarly, the copyright owner’s estate may be narrowed to account for valuable expressive uses that do not do much harm to the reproduction and distribution rights meant to be reserved to the owner. Acknowledgement of copyright’s conceptual underpinnings

---

370 I do not mean to assert that many or most cultural works do not borrow extensively from previous forms and texts, but just to suggest that gradations in originality and obviousness of borrowing exist.

371 See supra notes 123–124 and accompanying text.

372 BLACKSTONE, supra note 137, at *18.

373 See United States v. Causby, 328 U.S. 256, 261 (1946).
thus allows for compromise between the competing economic and free speech models of secondary use.
RISKING THE EIGHTH AMENDMENT: ARBITRARINESS, JURIES, AND DISCRETION IN CAPITAL CASES

JANET C. HOEFFEL*

Abstract: This Article argues that the stalled dialogue over the U.S. Supreme Court’s administration of capital punishment suffers from a fundamental misunderstanding of the first principles of the Eighth Amendment. Although the Court in Furman v. Georgia articulated an Eighth Amendment substantive right against the arbitrary imposition of death sentences, the Court later recast Furman to require procedures that merely reduced a substantial risk of arbitrariness. Instead, Furman mandates procedures that expose arbitrariness. The best vehicle for this is a review of jurors’ reasons for imposing death in an individual case. Although there are political and practical hurdles to mining the jurors’ reasons for imposing death, they are far from insurmountable. Absent a moratorium, this Article advocates change that informs and exposes the process of death.

As to impossibility, all I can say is that nothing is more true of [the legal] profession than that the most eminent among them, for 100 years, have testified with complete confidence that something is impossible which, once it is introduced, is found to be very easy of administration.

—F. Frankfurter

Introduction

We are at an impasse in the dialogue on the jurisprudence of capital punishment. In order to move forward, both proponents and opponents need to revisit the basic Eighth Amendment principle

* ©2005, Janet C. Hoeffel, Associate Professor of Law, Tulane Law School; J.D., Stanford Law School. I wish to thank Robert Weisberg, Victor Streib, Lloyd Bonfield, Michael Collins, Pamela Metzger, Felice Batlan, and Steve Singer for their helpful comments on earlier drafts of this article. I am grateful to my research assistants Scott Sherman, Rebekah Page, Jason Totoiu, and Andrew Cantor for their hard work. I also thank Dean Lawrence Ponoroff for the generous research grants that made this Article possible.

against the arbitrary imposition of the death penalty. Both sides mistakenly assume that there is no alternative to the U.S. Supreme Court’s current interpretation of its Eighth Amendment role as one of risk management.

In 1976 in *Gregg v. Georgia*, the Supreme Court held that procedures that substantially reduce the risk of arbitrariness in the imposition of the death penalty satisfy the Eighth Amendment. After instituting this risk-management system, we have never looked back. Looking back, however, both reveals the fundamental unacceptability of this system and suggests the alternative.

The Eighth Amendment command the Court purported to be fulfilling in *Gregg* came from *Furman v. Georgia*, decided four years earlier. *Furman* stood for the simple proposition that arbitrary imposition of death violated the Eighth Amendment. That substantive Eighth Amendment mandate against arbitrary decision making cannot be met by policing procedures that merely reduce the risk of arbitrariness. Under such a system, arbitrary decision making is tolerated.

One of the likely reasons the Court adopted risk-reduction procedures in *Gregg* is that the alternative—risk exposure—was unthinkable. To take *Furman* at face value is to require a searching review of direct evidence of arbitrariness, and direct evidence can only come from the jurors themselves. The jury, however, is ostensibly guarded by a wall of secrecy. It is well past time to scrutinize the propriety of jury secrecy in capital cases. Study of the concerns about revealing jurors’ thought processes uncovers, at bottom, the unspoken fear that revelation will expose arbitrariness. The demands of the Eighth Amendment require this exposure.

This Article breaks the impasse in the dialogue about capital punishment by changing the conversation from one about the adequacies of procedures used to reduce the risk of arbitrariness to one about the best procedures to expose arbitrariness. The imperfection in a

---

2 428 U.S. 153, 195 (1976) (opinion of Stewart, Powell, and Stevens, JJ.); see also id. at 188 (“Because of the uniqueness of the death penalty, *Furman v. Georgia* held that it could not be imposed under sentencing procedures that created a substantial risk that it would be inflicted in an arbitrary and capricious manner.”).
3 408 U.S. 238 (1972) (per curiam).
4 See infra notes 13–33 and accompanying text.
5 The evidence that we do have shows that arbitrary decision making does exist. See infra notes 100–06 and accompanying text.
6 The term “jury secrecy” in this Article refers to all procedures attendant to the non-disclosure of jury deliberations, both during trial and post-trial.
system that allows a risk of arbitrary imposition of death has been common ground between both proponents and opponents. The imperfection is either accepted as inevitable or rejected as inevitable. This proposal is targeted at the imperfection and the inevitability.

The Eighth Amendment proscription against arbitrary decision making can draw its meaning from a comparison to administrative law. Agency decisions are reviewed for arbitrariness as a matter of course, and the making of a record of the decisionmaker’s reasons is crucial to a searching review. In capital cases, the Supreme Court has openly struggled with the lack of a record in weighing the effects of error on the jury’s sentencing decision. The Court’s majority and dissenting opinions guess at whether the jurors thought about imposing the death penalty in a way that would have made their decision arbitrary. The Eighth Amendment mandate requires a procedure for uncovering, to the best possible degree, the jurors’ true reasons for their decision to impose death.

Part I of this Article demonstrates that, in Furman v. Georgia, the Court set forth the Eighth Amendment command as a substantive right against arbitrary imposition of death, but that, four years later in Gregg v. Georgia, the Court stepped back by declaring that Furman requires procedures that only reduce a substantial risk of arbitrariness.7 Part II then demonstrates how the standard critique of the Court’s administration of the death penalty, while helpful in its own right, has lost the true meaning of Furman and therefore lost the opportunity to move us forward.8 Part III discusses the content of arbitrary decision making in capital cases and establishes the need for a record of the jurors’ reasons in order to expose this arbitrariness.9 Part IV then studies the barriers to reviewing jury decision making. Enforcement of procedures supporting the tradition proves to be spotty and inconsistent, and the rationales underlying jury secrecy are surprisingly insubstantial when weighed against the Eighth Amendment mandate.10 Finally, Part V proposes mechanisms through which jurors’ reasons for imposing death may be reviewed. After discussing alternatives, this Article endorses post-trial interviews of jurors as the most reliable method for gathering evidence of arbitrariness.11

7 See infra notes 13–48 and accompanying text.
8 See infra notes 49–86 and accompanying text.
9 See infra notes 87–129 and accompanying text.
10 See infra notes 130–214 and accompanying text.
11 See infra notes 215–34 and accompanying text.
Although it may be that the best way to eliminate arbitrariness in capital cases is to suspend the death penalty, in the meantime, juries continue to sentence people to death. It is thirty years past time to review the process for imposing death: “If we would guide by the light of reason, we must let our minds be bold.”

I. Furman as Substantive Right, Gregg as Procedural Wrong

The Supreme Court’s shift in emphasis from Furman v. Georgia to Gregg v. Georgia four years later was an unfortunate, but foreseeable, concession to practicalities. If the Court interpreted Furman as it should have—as holding that the Eighth Amendment absolutely prohibited the arbitrary imposition of death—then how would there ever be proof of an arbitrary decision? It would be virtually impossible to prove that a jury acted arbitrarily because there is no record of why a jury decides to impose death. This raises the question asked in this Article: must that be so? Without asking that question, or deftly side-stepping it, judges and commentators have succumbed to the belief that, while not ideal, the most that could be done is to establish procedures that reduce the risk of an arbitrary result.

The starting premise is that Furman did in fact hold that the Eighth Amendment prohibited arbitrary imposition of death. At over 230 pages, the longest opinion ever written by the Court, with nine separate opinions, Furman can too easily be dispensed with as standing for no more than the narrow proposition in the per curiam opinion. The one-paragraph opinion held simply that “the imposition and carrying out of the death penalty in these [three] cases constitute cruel and unusual punishment in violation of the Eighth and Fourteenth Amendments.” However, a unifying principle of Furman can be loosely ascertained from the separately written opinions of the concurring Justices. The core commonality of the concurring opin-

14 408 U.S. at 239–40 (per curiam). The original ambition of some members of the Court was to decide whether the death penalty was “cruel and unusual punishment[]” prohibited by the Eighth Amendment, an issue the Court had avoided until Furman. See William J. Brennan, Jr., Constitutional Adjudication and the Death Penalty: A View from the Court, 100 Harv. L. Rev. 313, 321 (1986) (“We were clearly itching toward resolving [the issue.]”). As Justice Brennan recalled fourteen years after Furman, it was “clear that the difficult issue for everyone was how the Court could responsibly interpret the broadly worded prohibition against ‘cruel and unusual punishments.’” Id. at 322–23.
ions\textsuperscript{15} was the underlying concern that the jurors had made their decisions to impose death arbitrarily.\textsuperscript{16}

Justice Stewart’s short nine-paragraph concurring opinion is most often cited for the meaning of \textit{Furman}.\textsuperscript{17} Justice Stewart reasoned that because the death penalty was imposed so infrequently, the choice to do so in those few cases and not others was “cruel and unusual in the same way that being struck by lightning is cruel and unusual.”\textsuperscript{18} Petitioners were among “a capriciously selected random handful” who were chosen to die.\textsuperscript{19} Justice Stewart went no further than to “simply conclude that the Eighth and Fourteenth Amendments cannot tolerate the infliction of a sentence of death under legal systems that permit this unique penalty to be so wantonly and freakishly imposed.”\textsuperscript{20}

For Justice White, as for Justice Stewart, infrequency of imposition was the core problem. The infrequency meant that the death penalty could not be justified under societal goals of retribution or deterrence and therefore violated the Eighth Amendment.\textsuperscript{21} The infrequency of the imposition of death for even “the most atrocious crimes” meant “there is no meaningful basis for distinguishing the few cases in which it is imposed from the many cases in which it is not.”\textsuperscript{22}

\textsuperscript{15} Here, “the concurring opinions” refers to those of Justices Stewart, White, and Douglas, who found the operation of the death penalty unconstitutional in these cases. Justices Brennan and Marshall concurred in the result but concluded that the death penalty per se violated the prohibition against “cruel and unusual punishments.” See 408 U.S. at 257–306 (Brennan, J., concurring); \textit{id.} at 314–71 (Marshall, J., concurring).

\textsuperscript{16} Cf. Robert Weisberg, \textit{Deregulating Death}, 1983 \textit{Sup. Ct. Rev.} 305, 317 (“In the manner of literary criticism, one can extract unifying ‘themes’ in the \textit{Furman} opinions, such as the dangers of arbitrariness and discrimination . . . . But . . . there really is no doctrinal holding in \textit{Furman} . . . .”).

\textsuperscript{17} 408 U.S. at 306–14 (Stewart, J., concurring). This is likely because Justice Stewart’s opinion, short and concise, represented the least common denominator among the concurring opinions, and he was the only member of the Court in the majority in both \textit{Furman} and \textit{Gregg}, and wrote the plurality judgment opinion in \textit{Gregg}, which purported to interpret \textit{Furman}. See \textit{id.}

\textsuperscript{18} \textit{Id.} at 309.

\textsuperscript{19} \textit{Id.} at 309–10. Justice Stewart added, “My concurring Brothers have demonstrated that, if any basis can be discerned for the selection of these few to be sentenced to die, it is the constitutionally impermissible basis of race. But racial discrimination has not been proved, and I put it to one side.” \textit{Id.} at 310 (citation omitted).

\textsuperscript{20} \textit{Id.} Although Justice Stewart did not specifically use the word “arbitrary,” he subsequently used the word in \textit{Gregg} when stating the holding of \textit{Furman}. See \textit{id.; see also Gregg}, 428 U.S. at 188 (opinion of Stewart, J.) (using the words “arbitrary and capricious”).

\textsuperscript{21} \textit{Furman}, 408 U.S. at 311–12 (White, J., concurring).

\textsuperscript{22} \textit{Id.} at 313.
At bottom, Justice White’s concern was that the jury was arbitrarily withholding the death penalty.\textsuperscript{23}

Justice Douglas’s conclusion was that the death penalty statutes at issue, which gave jurors complete discretion to decide between life or death, allowed prejudice and discrimination to operate, and “discrimination is an ingredient not compatible with the idea of equal protection of the laws that is implicit in the ban on ‘cruel and unusual’ punishments.”\textsuperscript{24} Because “[p]eople live or die, dependent on the whim of one man or of 12,”\textsuperscript{25} a man may well be put to death “if he is poor and despised, and lacking political clout, or if he is a member of a suspect or unpopular minority.”\textsuperscript{26} Justice Douglas’s focus on an equal protection theme he found implicit in the Eighth Amendment was simply another form of the principle that “‘[a] penalty . . . should be considered ‘unusually’ imposed if it is administered arbitrarily or discriminatorily.’”\textsuperscript{27}

\textit{Furman} can most simply and directly be said to have held that the Eighth Amendment prohibits arbitrary imposition of the death penalty.\textsuperscript{28} Seen in this irreducible, straightforward manner, \textit{Furman} joined a line of rare but venerable cases, from \textit{Weems v. United States}\textsuperscript{29} to \textit{Trop}.
v. Dulles\textsuperscript{30} to Robinson v. California,\textsuperscript{31} which defined the substance of “cruel and unusual punishments” as encompassing more than acts of brutality and torture.\textsuperscript{32} The Eighth Amendment’s ban on “cruel and unusual punishments” also prohibited punishment imposed unnecessarily, disproportionately, excessively, and now, after Furman, death imposed arbitrarily.\textsuperscript{33}

This was not how Justice Stewart saw Furman four years later in Gregg v. Georgia.\textsuperscript{34} Rather, Justice Stewart, writing an opinion and issuing the plurality judgment of the Court, stated, “Because of the uniqueness of the death penalty, Furman held that it could not be imposed under sentencing procedures that created a substantial risk that it would be inflicted in an arbitrary and capricious manner.”\textsuperscript{35} Crucially, Furman’s holding also became the affirmative version of Justice Stewart’s rephrasing: the death penalty could be imposed under sentencing procedures that allowed for a risk, albeit something less than “substantial,” that it would be inflicted in an arbitrary and capricious manner.\textsuperscript{36} Gregg, through its reincarnation of Furman, then, could claim, “Furman mandates . . . that discretion [in a capital case] must be suitably directed and limited so as to minimize the risk of wholly arbitrary and capricious action.”\textsuperscript{37}

With these words, a system of risk-management was born.

Yet, neither “sentencing procedures” nor “risks” had been a part of Furman’s mandate. The concurring Justices in Furman had identified the results, and not the presence or absence of procedures,

\begin{itemize}
    \item \textsuperscript{30} 356 U.S. 86, 99–103 (1958) (holding expatriation cruel and unusual).
    \item \textsuperscript{31} 370 U.S. 660, 667 (1962) (holding imprisonment for narcotics addiction cruel and unusual).
    \item \textsuperscript{32} See Weems, 217 U.S. at 372 (“[S]urely [the Framers] intended more than to register a fear of the forms of abuse that went out of practice with the Stuarts.”).
    \item \textsuperscript{33} At least one scholar has suggested that a prohibition against arbitrary infliction of a severe punishment was inherent in the history of the Cruel and Unusual Clause. See Anthony F. Granucci, “Nor Cruel and Unusual Punishments Inicted:” The Original Meaning, 57 Cal. L. Rev. 839, 857–60 (1969).
    \item \textsuperscript{34} 428 U.S. at 188 (opinion of Stewart, Powell, and Stevens, J J.).
    \item \textsuperscript{35} Id. (opinion of Stewart, Powell, and Stevens, J J.) (emphasis added).
    \item \textsuperscript{36} See id. (opinion of Stewart, Powell, and Stevens, J J.).
    \item \textsuperscript{37} Id. at 189 (opinion of Stewart, Powell, and Stevens, J J.) (emphasis added). This “wholly arbitrary and capricious” phrasing became a favorite of the Court’s. See Arave v. Creech, 507 U.S. 463, 470 (1993) (quoting phrase); Lewis, 497 U.S. at 774 (same); Barclay v. Florida, 463 U.S. 939, 950 (1983) (same); Zant v. Stephens, 462 U.S. 862, 874 (1983) (same). Justice Marshall took issue with the use of “wholly” in Barclay. “This implies that in death cases there are degrees of acceptable arbitrariness and that there exists some undefined point at which a sentence crosses over into the nether world of ‘wholly’ arbitrary decisionmaking.” 463 U.S. at 987 (Marshall, J., dissenting).
\end{itemize}
as the primary evil. Justices Stewart and White had not even mentioned procedures—neither those procedures in place nor those that might be required.\textsuperscript{38} And no concurring Justice had settled merely for eliminating a “risk,” because they each had simply assumed that arbitrary imposition had, in fact, occurred.\textsuperscript{39} Although one might argue that procedures to manage risk were what the concurring Justices had in mind, that was not what they said.

In characterizing the holding of \textit{Furman} in \textit{Gregg}, Justice Stewart was faced with two realities. First, if he recognized \textit{Furman} as outlawing arbitrary imposition of death, then any statutory scheme approved would have to provide a means for empirical proof of arbitrariness.\textsuperscript{40} That was a subject that had not been broached; even the concurring Justices in \textit{Furman} conceded that they were proceeding under an assumption of arbitrariness rather than an empirical foundation. The only method for providing an empirical foundation

\textsuperscript{38} In addition, “Justice Douglas essentially challenges the states to resolve his social and political discomfort over the \textit{products} of the death penalty.” Weisberg, \textit{supra} note 16, at 316 (emphasis added).

\textsuperscript{39} Because Justice Stewart in \textit{Furman} had no actual proof of why the jurors had decided to impose the penalty in the very few cases in which they had done so, he used the infrequency as a proxy for proof. \textit{See} 408 U.S. at 309 (Stewart, J., concurring). Justice White explicitly admitted his lack of proof and relied upon his own personal experience to tell him that jurors were acting arbitrarily. \textit{See id.} at 313 (White, J., concurring) (“I need not restate the facts and figures that appear in the opinions of my Brethren. Nor can I ‘prove’ my conclusion from these data. But, like my Brethren, I must arrive at judgment; and I can do no more than state a conclusion based on 10 years of almost daily exposure to the facts and circumstances of hundreds and hundreds of federal and state criminal cases involving crimes for which death is the authorized penalty.”). Justice Douglas did amass some proof that jurors acted with discrimination. \textit{Id.} at 249–52 (Douglas, J., concurring) (citing study of capital cases in Texas and comments of former warden of Sing Sing Prison and former Attorney General Ramsey Clark).

\textsuperscript{40} In a Note published in the \textit{Harvard Law Review} after \textit{Furman} and before \textit{Gregg}, the author suggested that the Court could not interpret \textit{Furman} as an “arbitrariness-in-fact” test because doing so would require an empirical foundation, which conceded the concurring Justices did not have. Note, \textit{Discretion and the Constitutionality of the New Death Penalty Statutes}, 87 \textit{Harv. L. Rev.} 1690, 1694 (1974). To interpret \textit{Furman} instead as condemning unfettered discretion in capital sentencing schemes would justify the Court’s having acted in \textit{Furman} and companion cases without any sound empirical foundation, and would avoid the need for judicial speculation about how statutes will be administered. Moreover, . . . requiring the state to restrict discretion may be the only feasible method to assure re-duction of the arbitrariness which three of the concurring Justices found prescribed by the eighth amendment. Such a requirement would also limit official complicity in whatever arbitrariness is not eliminated.

\textit{Id.} at 1695–96.
would be to pierce the veil of jury secrecy, and that suggestion was, perhaps, unimaginable.

The second practical reality was that the states had responded to *Furman* with overwhelming support for capital punishment.\(^{41}\) Anxious to meet the demand of their constituents, the state legislatures adopted the most expedient mechanism at their disposal. Guided discretion statutes had already been contemplated and discussed. In 1959, the American Law Institute suggested a version of such a statute in the Model Penal Code. The Code provided a list of aggravating circumstances and a list of mitigating circumstances that the sentencer could consider, none of which were exclusive, and instructed that the penalty of death should not be imposed unless the sentencer found that one of the enumerated aggravating circumstances existed and there were no mitigating circumstances sufficiently substantial to call for mercy.\(^{42}\) Guided discretion statutes such as this were already in circulation when the Court decided *Furman*.

Given the lack of perceived alternatives to guided discretion statutes, Justice Stewart accepted guided discretion as the answer to *Furman*.\(^{43}\) In order to hold that guided discretion statutes complied with *Furman*, Justice Stewart had to find that *Furman* commanded only risk-managing procedures. Guided discretion statutes allow ample room for arbitrary decision making. In fact, six members of the Court said as much one year before *Furman*. Justice Harlan, writing for the Court in *McGautha v. California*,\(^{44}\) considered and rejected the Model Penal Code’s suggested guidelines. In holding that unfettered discretion to impose death was not unconstitutional, he argued that the expression of standards was useless to guard against the supposed evils of discre-

\(^{41}\) In the year following *Furman*, more than half the state legislatures nationwide introduced bills to restore capital punishment. See *Capital Punishment in the United States: A Documentary History* 148 (Bryan Vila & Cynthia Morris, eds., 1997). Within two years of *Furman*, over half of the states had passed new death penalty legislation. See Note, *supra* note 40, at 1691 & n.6 (citing statutes). Public support for the death penalty increased after *Furman*, from 53% in 1972, to 60% in 1976. Samuel R. Gross, *Still Unfair, Still Arbitrary—But Do We Care?*, 26 Ohio N.U.L. Rev. 517, 521 (2000).


\(^{43}\) The seeds of *Gregg’s* solution of guided discretion were sewn both in *McGautha* and in Chief Justice Burger’s dissent in *Furman*. Chief Justice Burger’s dissent essentially gave the states their marching orders: he offered that the concurring Justices’ concern about arbitrary punishment could be met “by providing standards for juries and judges to follow in determining the sentence in capital cases or by more narrowly defining the crimes for which the penalty is to be imposed.” 408 U.S. at 400 (Burger, C.J., dissenting).

\(^{44}\) 402 U.S. at 207.
tionless sentencing: “[S]uch criteria do not purport to provide more than the most minimal control over the sentencing authority’s exercise of discretion . . . . [a]nd, of course, they provide no protection against the jury determined to decide on whimsy or caprice.”

With a flourish of the pen, guided discretion statutes were imported into the substance of the Eighth Amendment. The transformation of *Furman*’s substantive mandate into a procedural due process demand for “guided discretion” was complete. The only matter left to the courts was to monitor the level of risk tolerable under the Constitution. So began the states’ and the Court’s long and lumbering engagement with guided discretion statutes and the problems foreshadowed by Justice Harlan. The standard critique has correctly decried this ill-fated engagement. However, if *Furman* really meant what it said, then the critique should have gone much further.

II. THE DEAD ENDS OF THE STANDARD CRITIQUE

The limitation of the standard critique of the Court’s administration of the death penalty lies in the fact that it does not question *Gregg v. Georgia*’s retooling of *Furman v. Georgia* from an opinion intolerant of arbitrariness to one tolerant of a risk. Rather, the critique focuses on the fact that the Court’s administration of this risk-management system has failed. This is valid criticism in its own right, but it can only lead to dead ends. The criticism has lead to one of two places: a call for either the end of the system or a better risk-management system.

---

45 Id. Justice Harlan’s critique of those guidelines would be echoed again and again over the next thirty years. Robert Weisberg wrote that Justice Harlan’s statement in *McGautha* “was the Court’s (soon unheeded) warning to itself that the entire enterprise of subjecting capital punishment to legal rules was hopeless and unnecessary.” Weisberg, supra note 16, at 308.

46 See *Furman*, 408 U.S. at 399 (Burger, C.J., dissenting) (arguing that the concurring opinions called for guided discretion, which was “essentially and exclusively a procedural due process argument”).

47 See *Turner v. Murray*, 476 U.S. 28, 36 n.8 (1986) (“[T]he only question is at what point that risk becomes constitutionally unacceptable.”).

48 See, e.g., *Callins v. Collins*, 510 U.S. 1141, 1145 (1994) (Blackmun, J., dissenting from denial of certiorari) (“[T]he Court has chosen to deregulate the entire enterprise, replacing, it would seem, substantive constitutional requirements with mere aesthetics”).


50 See, e.g., *Callins v. Collins*, 510 U.S. 1141, 1145 (1994) (Blackmun, J., dissenting from denial of certiorari) (“[T]he death penalty experiment has failed. It is virtually self-evident to me now that no combination of procedural rules or substantive regulations ever can save the death penalty from its inherent constitutional deficiencies.”).
The former will not be compelled any time soon, and the latter unfortunately continues to accept risk management over risk elimination.

The conventional critique describes two problems in the risk-management scheme. First, the Court has not demanded much “guidance” in guided discretion, therefore establishing the level of risk unacceptably high. Second, the principles behind guided discretion are in tension with a separate Eighth Amendment principle of individuality, which calls for the unlimited discretion of the jury to consider mitigating evidence on behalf of the defendant. However, if *Furman* is viewed as a substantive command against arbitrary impositions of death, the lack of guidance problem, while real, misses the mark, and the dueling principles problem is nonexistent.

A. The Lack of Guidance Problem

Critics have ably demonstrated that the Supreme Court essentially abandoned a regime of meaningful guidance. Two decades ago, Robert Weisberg wrote that “the Court has reduced the law of the penalty trial to almost a bare aesthetic exhortation that the states just do something—anything—to give the penalty trial a legal appearance.”51 The Court has approved statutory schemes that do no more than simply narrow the class of offenders eligible for the death penalty, allowing unguided discretion to reign at that point.52 The Court

---

52 In *Zant v. Stephens*, the Supreme Court affirmed Georgia’s statutory scheme where “the aggravating circumstance merely performs the function of narrowing the category of persons convicted of murder who are eligible for the death penalty,” 462 U.S. 862, 875 (1983). The approved statute gave the jury “absolute discretion” to decide whether to impose life or death after finding the aggravating circumstance. *Id.* at 871. As Carol and Jordan Steiker observed:

If we are worried that the failure to provide precise guidance to capital sentencers may lead them to use irrelevant characteristics (like physical attractiveness) or impermissible ones (like race or class) to determine who should live and who should die from among the equally eligible, this problem is not resolved merely by narrowing the range of persons among whom the sentencer can discriminate.

has also upheld vague aggravating factors\textsuperscript{53} and eliminated mandatory proportionality review of sentences,\textsuperscript{54} further diminishing protection against arbitrary decision making.

Critics’ suggestions for remediying these problems, however, are unsatisfactory. Some have suggested reworking those same procedures,\textsuperscript{55} but guided discretion procedures, no matter how refined, entail very real risks of arbitrary decisions to impose death. Others have ultimately concluded that it is impossible to guide the inevitably discretionary decision whether to impose death,\textsuperscript{56} and instead focus mockery of this Court’s precedents concerning capital sentencing procedures.”) (quoting \textit{Gregg}, 428 U.S. at 195 n.47 (opinion of Stewart, Powell, and Stevens, JJ.)).

\textsuperscript{53} See, e.g., \textit{Tuilaepa v. California}, 512 U.S. 967, 978 (1994) (approving “circumstances of the crime” as an aggravating circumstance in California); \textit{Arave v. Creech}, 507 U.S. 463, 471–72 (1993) (upholding Idaho’s “utter disregard for human life” circumstance with the interpretation that the defendant be a “cold-blooded, pitiless slayer”); \textit{Walton v. Arizona}, 497 U.S. 639, 654–55 (1990) (upholding Arizona’s “especially heinous, cruel or depraved” circumstance as defined by the Arizona Supreme Court as when the perpetrator “relishes the murder, evidencing debasement or perversion,” or ‘shows an indifference to the suffering of the victim and evidences a sense of pleasure’ in the killing”); \textit{Profitt v. Florida}, 428 U.S. 242, 255–56 (1976) (upholding the “especially heinous, atrocious or cruel” circumstance on Florida’s restriction to include only “‘the conscienceless or pitiless crime which is unnecessarily torturous to the victim’”). Justice Blackmun’s dissents in \textit{Walton v. Arizona} and \textit{Tuilaepa v. California} roundly criticized the Court’s conclusions that there were any limiting circumstances at all to these factors, in principle or in practice. \textit{Tuilaepa}, 512 U.S. at 984–96 (Blackmun, J., dissenting); \textit{Walton}, 497 U.S. at 692–99 (Blackmun, J., dissenting); see also Richard A. Rosen, \textit{The “Especially Heinous” Aggravating Circumstance in Capital Cases—The Standardless Standard}, 64 N.C. L. Rev. 941, 942 (1986) (conducting state-by-state analysis of appellate decisions and concluding that their inconsistent and overbroad application of the “especially heinous” circumstance does not guide discretion); Steiker & Steiker, \textit{Sober Second Thoughts}, supra note 52, at 373–74 (arguing that approval of vague aggravating circumstances and lack of limits on the number of aggravating factors does not narrow the choice of who dies).

\textsuperscript{54} Pulley v. Harris, 465 U.S. 37, 50–51 (1984). In dissent in \textit{Pulley}, Justice Brennan voiced his dissatisfaction with the current state of affairs:

The results obtained by many States that undertake such proportionality review, pursuant to either state statute or judicial decision, convince me that this form of appellate review serves to eliminate some, if only a small part, of the irrationality that infects the current imposition of death sentences throughout the various States.

\textit{Id.} at 67–68 (Brennan, J., dissenting).


\textsuperscript{56} See \textit{Callins}, 510 U.S. at 1153 (Blackmun, J., dissenting from denial of certiorari) (arguing that “the decision whether a human being should live or die is so inherently subjec-
on points in the system other than the decision-making process. To avoid a hard look at the jury’s decision, however, is to escape Furman’s constitutional mandate. As long as there is discretion, there is the possibility of prejudice, mistake, and caprice. And as long as there is prejudice, mistake, and caprice, it is the constitutional obligation of the courts to expose it, not avoid it.

Members of the Court have readily acknowledged that arbitrary imposition of death will occur within its system of managing risk. Then-Justice Rehnquist recognized the inevitability of arbitrariness in Woodson v. North Carolina. In discussing the meaninglessness of a proportionality review of a death sentence, he said:

The plurality seems to believe that provision for appellate review will afford a check upon the instances of juror arbitrariness in a discretionary system. But it is not at all apparent that appellate review of death sentences, through a productive—rife with all of life’s understandings, experiences, prejudices, and passions—that it inevitably defies the rationality and consistency required by the Constitution.”); Patrick E. Higginbotham, Juries and the Death Penalty, 41 Case W. Res. L. Rev. 1047, 1048–49 (1991) (”[T]he ultimate call is visceral. The decision must occur past the point to which legalistic reasoning can carry; it necessarily reflects a gutlevel hunch as to what is just.”); Steiker & Steiker, Sober Second Thoughts, supra note 52, at 413 (”If the capital punishment decision can be neither ‘true’ nor ‘false’ (as facts are said to be) nor ‘correct’ or ‘incorrect’ (as legal judgments are said to be), but instead represents ‘an existential moment of moral perception,’ then refining the directions given to the decisionmakers (the jurors) cannot assist, but rather will merely obscure the fundamentally moral choice that they must make.” (citations omitted)); Weisberg, supra note 16, at 395 (arguing that the ineffective procedures built around the moment of decision “ha[ve] enabled us to avoid acknowledging the inevitably unsystematic, irreducibly personal moral elements of the choice to administer the death penalty”).

57 See Weisberg, supra note 16, at 359–60 (arguing that after the Court’s deregulation of the administration of capital punishment, we “have to focus on other points in the system of death penalty decisionmaking other than the moment of decision by the sentencer”). The proposals include: establishing narrowing factors which effectively reduce the kind and number of homicides for which the jury can impose death, Scott W. Howe, The Failed Case for Eighth Amendment Regulation of the Capital-Sentencing Trial, 146 U. Pa. L. Rev. 795, 848 (1998); Steiker & Steiker, Sober Second Thoughts, supra note 52 at 415–17; making proportionality decisions to reduce further who can be considered for the death penalty, Howe, supra note 52, at 848–49; Steiker & Steiker, Sober Second Thoughts, supra note 52, at 417–18; mandating quality of counsel for capital defendants, Howe, supra note 52, at 856–57; Steiker & Steiker, Sober Second Thoughts, supra note 52, at 421–23; refining rigorous appellate review, Weisberg, supra note 16, at 359–60; expanding post-conviction opportunities, Steiker & Steiker, Sober Second Thoughts, supra note 52, at 423–25; and scrutinizing prosecutorial discretion, see Hans Zeisel, Race Bias in the Administration of the Death Penalty: The Florida Experience, 95 Harv. L. Rev. 456, 466–68 (1981) (focusing on prosecutorial discretion as a source of racial bias).

cess of comparing the facts of one case in which a death sentence was imposed with the facts of another in which such a sentence was imposed, will afford any meaningful protection against whatever arbitrariness results from jury discretion.59

The same can also be said of the individual “arbitrariness” review approved by the Court in Gregg. Georgia’s procedure for reviewing arbitrariness in the jury’s imposition of death involved a review of the trial judge’s answer to a questionnaire, which asked whether passion, prejudice, bias, or any other arbitrary factor, including race, influenced the jury’s decision.60 It is impossible to imagine this was anything other than a symbolic gesture. Excepting the rare occasion where jury misconduct came to light during the trial, this was likely a perfunctory check in the “No” column by the trial judge.

Justice Powell likewise embraced arbitrariness as inevitable in drafting the Court’s decision in McCleskey v. Kemp.61 McCleskey is the low watermark of the Court’s operation of a system that monitors an acceptable level of risk of arbitrary imposition of death. The defense presented a statistical study showing that jurors in Georgia were making decisions to impose death based on the defendant’s and the victim’s race.62 Justice Powell, writing for a bare majority of the Court, assumed the accuracy of the statistics but held that the study did not show an Eighth Amendment violation.63 Justice Powell’s opinion reads like an apology for the system. He claimed that, despite the fact that the process “‘has its weaknesses and the potential for misuse’”64 and “‘there can be ‘no perfect procedure for deciding in which cases

59 Id. at 316 (Rehnquist, J., dissenting) (internal citation omitted). State proportionality review has proven largely ineffective. See Leigh B. Bienen, The Proportionality Review of Capital Cases by State High Courts After Gregg: Only “The Appearance of Justice”? 87 J. CRIM. L. & CRIMINOLOGY 130, 133 (1996) (“[T]he majority of state high courts reduced proportionality review to a perfunctory exercise.”).

60 See Gregg, 428 U.S. at 211–12 (White, J., concurring) (describing questionnaire). Georgia still has such a form, which contains the question, “Was the jury impermissibly influenced by passion, prejudice, or any other arbitrary factor when imposing sentence?” followed by the choice of checking, “Yes” or “No,” and then “If the answer is yes, explain:” followed by two and one-half lines. See Supreme Court of Georgia, Report of the Trial Judge, http://www2.state.ga.us/Courts/Supreme/rules_UAP/uasect6.htm (last visited Aug. 24, 2005).


62 The “Baldus study” indicated that “the jury more likely than not would have spared McCleskey’s life had his victim been black” and that blacks who kill whites are sentenced to death “at nearly 22 times the rate of blacks who kill blacks, and more than 7 times the rate of whites who kill blacks.” Id. at 325, 327 (Brennan, J., dissenting).

63 Id. at 308–13.

64 Id. at 313 (quoting Singer v. United States, 380 U.S. 24, 35 (1965)).
governmental authority should be used to impose death,’” constitutional guarantees are met when the sentencer’s decision is “surrounded with safeguards to make it as fair as possible.” Therefore, despite the fact that there was a high probability that the jurors in McCleskey’s case imposed death based on the arbitrary factor of race, once the risk had been reduced as far as the Court deemed was practicable, the Eighth Amendment demand had been met.

To the dissent’s claim that the statistics demonstrated a capital punishment system that lacked the “uniquely high degree of rationality” called for in capital cases, Justice Powell defensively retorted that the dissent gave “no suggestion . . . as to how greater ‘rationality’ could be achieved under any type of statute that authorizes capital punishment” and that the “dissent’s call for greater rationality is no less than a claim that a capital punishment system cannot be administered in accord with the Constitution.” Hence, the impasse in the dialogue: either accept a certain level of risk or abandon the project.

"Id. (quoting Zant, 462 U.S. at 884 (quoting Lockett v. Ohio, 438 U.S. 586, 605 (1978) (plurality opinion))). “[N]o perfect procedure” became one of the Court’s pet phrases. See, e.g., Pulley, 465 U.S. at 54 (quoting phrase); Zant, 462 U.S. at 884 (same).

66 McCleskey, 481 U.S. at 313 (quoting Singer, 380 U.S. at 35). Justice Blackmun would later challenge that “the majority turned its back on McCleskey’s claims, apparently troubled by the fact that Georgia had instituted more procedural and substantive safeguards than most other States since Furman, but was still unable to stamp out the virus of racism.” Callins, 510 U.S. at 1154 (Blackmun, J., dissenting from denial of certiorari).

67 Justice Powell concluded that the statistics, while possibly showing a likelihood that race was a factor, did not create a constitutionally significant risk of racial bias affecting Georgia’s capital process. McCleskey, 481 U.S. at 308, 313. Three years after Justice Powell retired from the Court in 1987, he told his biographer that he would change his vote in McCleskey and in any other capital case, including Furman, where he voted to uphold capital punishment. See RANDALL COYNE & LYN ENTZEROTH, CAPITAL PUNISHMENT AND THE JUDICIAL PROCESS 202–03 (2d. ed. 2001) (citing JOHN CALVIN JEFFRIES, JUSTICE LEWIS F. POWELL, JR. 451–52 (1994)).

68 481 U.S. at 335 (Brennan, J., dissenting).

69 Id. at 313 n.37. The Court’s emphasis on procedures over results is exemplified by comparing Turner v. Murray, where the Court agreed that the risk of racial bias infecting a capital trial was so great that jurors could be questioned about it during voir dire, with McCleskey, where the Court ignored evidence of racial bias in jurors’ decisions to impose death. See generally McCleskey, 481 U.S. 279; Turner v. Murray, 476 U.S. 28 (1986). Discussing Turner and McCleskey, Albert Alschuler wrote, “The Court added one more gargoyle to the front end of the temple of justice while turning its eyes from back-end outcomes. The Court’s approach appeared to be, ‘Procedure yes, substance no.’” Albert W. Alschuler, The Supreme Court and the Jury: Voir Dire, Peremptory Challenges, and the Review of Jury Verdicts, 56 U. CHI. L. REV. 153, 229 (1989).

70 Steiker and Steiker note the tension between the right to jury trials and the ensuing discretion in capital punishment cases:
B. The Dueling Principles Problem

The second standard claim of critics of the Court’s administration of the death penalty is that *Furman*’s Eighth Amendment command, which is described as a command for standardization and consistency, is in irreconcilable conflict with the *Woodson/Lockett v. Ohio* 71 Eighth Amendment command to individualize the sentencing decision to the characteristics of the individual offenders. The perception of the tension has forced one of two decisions among some members of the Court: Justice Scalia has chosen to abandon individualized consideration, 72 while Justice Blackmun chose to abandon the death penalty altogether. 73 Academics who agree there is a tension either decide the death penalty does not survive the tension 74 or simply choose to make suggestions on points of the process other than the jury’s decision. 75

However, the concept of this “tension” misperceives *Furman*. If the goal of the system is an avoidance of arbitrary death in the manner *Furman* intended, individuality enhances the goal. The principles

---


71 See generally *Lockett*, 438 U.S. 586 (plurality opinion); *Woodson*, 428 U.S. 280.


73 *Callins*, 510 U.S. at 1145 (Blackmun, J., dissenting from denial of certiorari); see also *Godfrey v. Georgia*, 446 U.S. 420, 442 (1980) (Marshall, J., concurring) (“[T]he effort to eliminate arbitrariness in the infliction of that ultimate sanction is so plainly doomed to failure that—it and the death penalty—must be abandoned altogether.”).

74 See Steven G. Gey, *Justice Scalia’s Death Penalty*, 20 Fla. St. U. L. Rev. 67, 103 (1992) (“The Court’s two objectives are not, as Justice Scalia argues, irreconcilable with each other. Rather, they are irreconcilable with the death penalty.”); Margaret Jane Radin, *Cruel Punishment and Respect for Persons: Super Due Process for Death*, 53 S. Cal. L. Rev. 1143, 1155 (1980) (“[I]f death as a punishment requires both maximum flexibility and nonarbitrari ness, and these requirements cannot both be met, . . . then death cannot be a permissible punishment.”); see also Steiker & Steiker, *Let God Sort Them Out*, supra note 52, at 839–41, 868 (arguing individualization has constitutional grounding but leads to arbitrariness, meaning “*Furman* must be jettisoned, or *Furman*’s failure to fulfill its promise of principled, nonarbitrary decisionmaking renders the death penalty unconstitutional”).

75 For a discussion of critics’ suggestions, see supra note 57.
only conflict if nonarbitrariness is equated with standardization, which has as its goal ensuring consistency between punishments.\footnote{See, e.g., Howe, \textit{supra} note 57, at 810–11 (equating “consistency” with “nonarbitrariness” as the driving theory of the Court).} However, consistency was not, and has not been, the Court’s end goal.

The underlying concern of the concurring Justices in \textit{Furman} was the jury’s arbitrariness \textit{in an individual case}. Certainly there were phrases that sounded in arbitrariness \textit{between} cases. Justice White said of the death penalty, for example, that “there is no meaningful basis for distinguishing the few cases in which it is imposed from the many cases in which it is not,”\footnote{\textit{Furman}, 408 U.S. at 313 (White, J., concurring).} and Chief Justice Burger reiterated the concern of the concurring Justices as “not that too few have been sentenced to die, but that the selection process has followed no rational pattern.”\footnote{Id. at 399 (Burger, C.J., dissenting); see also \textit{id.} at 294 (Brennan, J., concurring) (“No one has yet suggested a rational basis that could differentiate in those terms the few who die from the many who go to prison.”).}

The concurring Justices, however, did not condemn simply the inconsistent pattern of imposition of death. Rather, the inconsistent pattern informed them that the jurors in the individual cases were acting arbitrarily in imposing the death penalty.\footnote{Justice Douglas offered that the arbitrary factors at work in individual cases were “race, religion, wealth, social position, or class.” \textit{Id.} at 242 (Douglas, J., concurring). Justice Stewart made a nod in that direction. \textit{Id.} at 310 (Stewart, J., concurring) (“[I]f any basis can be discerned for the selection of these few to be sentenced to die, it is the constitutionally impermissible basis of race.”).} Since jurors sat only in one case and were not repeat players, it could only be their lack of reasoned judgment in the individual case that produced the irregular pattern.\footnote{The Court in \textit{Gregg} also seemed to understand this individualist aspect of the meaning of “arbitrary and capricious.” The mandatory appellate review provided by Georgia’s death penalty statute, and given a ringing endorsement by the Court, had two parts. The first was a proportionality review—ensuring a pattern of rationality between cases according to the checklist. But the second was a determination “[w]hether the sentence of death was imposed under the influence of passion, prejudice, or any other arbitrary factor,” based partly on a six and one-half page questionnaire filled out by the trial judge that disclosed whether race played a role in the case. 428 U.S. at 166–67 (opinion of Stewart, Powell, and Stevens, J.J.) (describing questionnaire). This was clearly a review of whether arbitrariness was at work in the individual case. See \textit{Zant}, 462 U.S. at 890 (describing Georgia’s mandatory appellate review as designed both to “avoid arbitrariness and to assure proportionality”).}

Therefore, if arbitrariness is defined as an unprincipled or irrational imposition of punishment on a particular defendant, then individualized consideration not only does not conflict with it, but such
consideration also enhances the “accuracy”\textsuperscript{81} of the imposition of the punishment.\textsuperscript{82} Although accuracy in sentencing is a difficult concept given the amount of discretion and subjectivity involved, more information about the defendant and his circumstances allows a sentencer to make a better and more informed decision as to who deserves to die.\textsuperscript{83} As long as the jury uses the information in a rational, nonarbitrary way, then the jury is better able to function as the “conscience of the community”\textsuperscript{84} in deciding whether to exercise mercy in an individual case.\textsuperscript{85}

The standard critique accepts defeat too readily. It believes that jurors’ decisions to impose death are impervious to review and hence arbitrariness is inevitable. The defeat stems from a misperception of the constitutional mandate in death cases. The break in the impasse is to see that \textit{Furman} set the Eighth Amendment floor higher than \textit{Gregg}: intolerance of arbitrariness requires a system that exposes it, not one that hides it and guesses at its existence. Meeting \textit{Furman}’s command requires embracing the moment of decision, reviewing the jurors’ reasons for imposing death, and forcing confrontation of juror arbitrariness.\textsuperscript{86}

\textsuperscript{81} See \textit{Gilmore v. Taylor}, 508 U.S. 333, 342 (1993) (“[T]he Eighth Amendment requires a greater degree of accuracy . . . than would be true in a noncapital case.”).

\textsuperscript{82} See Scott E. Sundby, \textit{The Lockett Paradox: Reconciling Guided Discretion and Unguided Mitigation in Capital Sentencing}, 38 UCLA L. Rev. 1147, 1175–76 (1991) (arguing the two principles can be reconciled because both are “aimed at the same narrowing objective of identifying, as precisely as possible, who is within the state’s power to execute”).

\textsuperscript{83} Then-Justice Rehnquist, in his dissent in \textit{Lockett v. Ohio}, was concerned that the “more” would be “any fact, however bizarre” that the defendant wished to introduce, and would therefore “codify and institutionalize” arbitrariness. 438 U.S. at 631 (Rehnquist, J., dissenting). However, as Scott Sundby argues, there is no evidence that expanding the amount of mitigating evidence will reduce reliability, and no proof either that litigants will risk alienating the jury by introducing “bizarre” evidence or that jurors would react to it. See Sundby, \textit{supra} note 82, at 1182–83.


\textsuperscript{85} See also Radin, \textit{supra} note 74, at 1159 (“The substantive judgment to be made is a moral judgment: Does this person deserve death as punishment? The requirement that aggravating and mitigating factors be weighed is a requirement aimed at greater accuracy in making that moral judgment. The analysis assumes that it is theoretically possible for a person to deserve death as punishment, and that death is not morally ruled out in all cases. The analysis focuses on the moral requirements necessary to render noncruel the process of deciding that this person deserves death.” (citation omitted)).

\textsuperscript{86} At least one court has taken this challenge seriously. In \textit{Dobbs v. Zant}, the District Court for the Northern District of Georgia allowed a capital defendant to present juror testimony regarding racial bias in an effort to prove an “unacceptable risk” that race affected the sentencing decision. 720 F. Supp. 1566, 1574–79 (N.D. Ga. 1989).
III. Defining and Making a Record of Arbitrariness

There is a world of difference between procedures designed to reduce a risk of arbitrariness and procedures designed to expose arbitrary decision making. In order to discover arbitrariness, we must first define what constitutes an arbitrary imposition of death, and second, we must recognize that a record of the decision is necessary.

A. What Is Arbitrariness in Imposing Death?

When the Court in Gregg v. Georgia chose to use the phrase “arbitrary and capricious” to summarize the concurrences in Furman v. Georgia, it was not a stranger to the legal meaning of the phrase in other contexts. The phrase is talismanic in administrative law: An agency rule is deemed arbitrary and capricious if the agency has relied on factors which Congress has not intended it to consider, entirely failed to consider an important aspect of the problem, offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise.

Translated into jury sentencing terms, an arbitrary and capricious decision to impose death would include one where the jurors relied on factors that jurors should not consider—such as race, an aggravating factor subsequently invalidated by the Court, or the availability of appellate review to correct death sentences—or where jurors entirely failed to consider an important aspect, such as mitigating evidence.

89 See Furman, 408 U.S. at 242 (Douglas, J., concurring) (arguing that “unusual” punishment would include imposition of the penalty “by reason of [the defendant’s] race, religion, wealth, social position, or class” or penalties “imposed under a procedure that gives room for the play of such prejudices”).
son, for an illegitimate reason, or for no reason at all, it would be an arbitrary and capricious decision.93

The Court has occasionally pointed out examples of what it considered arbitrary decision making in capital cases. In *Mills v. Maryland* the Court found that, because of the instructions and the sentencing form given to the jury, there was a substantial probability that the jurors thought they were precluded from considering any mitigating evidence unless they unanimously agreed on the existence of particular mitigating circumstances.94 Justice Blackmun wrote for the Court that “it would certainly be the height of arbitrariness to allow or require the imposition of the death penalty under [such] circumstances.”95 In *Caldwell v. Mississippi*, the Court held it was reversible error for the prosecution to argue to the jury that the appellate courts, and not the jury, would have the final word on the appropriateness of the death penalty.96 Justice Marshall wrote for the Court that “for a sentencer to impose a death sentence out of a desire to avoid responsibility for its decision presents the spectre of the imposition of death based on a factor wholly irrelevant to legitimate sentencing concerns.”97 Justice O’Connor concurred, stating that the prosecution’s misinformation “creat[ed] an unacceptable risk that ‘the death penalty [may have been] meted out arbitrarily or capriciously,’98 or through ‘whim . . . or mistake.’”99

---

93 “Arbitrary and capricious” has also been considered by the Court in capital cases to include concepts such as “whim . . . or mistake,” *Caldwell*, 472 U.S. at 343 (O’Connor, J., concurring) (quoting *Eddings v. Oklahoma*, 455 U.S. 104, 118 (1982) (O’Connor, J., concurring)), “whim, caprice or prejudice,” *Callins v. Collins*, 510 U.S. 1141, 1144 (1994) (Blackmun, J., dissenting from denial of certiorari), and “caprice or emotion,” *Gardner v. Florida*, 430 U.S. 349, 358 (1977).

94 486 U.S. at 384.
95 *Id.* at 374.
96 472 U.S. at 336, 340–41.
97 *Id.* at 332.
98 *Id.* at 343 (O’Connor, J., concurring) (quoting *California v. Ramos*, 463 U.S. 992, 1020 (1983) (Marshall, J., dissenting) (alteration in original)).
99 *Id.* at 343 (O’Connor, J., concurring) (quoting *Eddings*, 455 U.S. at 118 (O’Connor, J., concurring)); see also *Ramos*, 463 U.S. at 1019–20 (Marshall, J., dissenting) (arguing that jury instruction informing jurors of governor’s power to commute a life sentence had jury guessing about commutation and parole, and “[s]entencing decisions based on such groundless predictions are clearly arbitrary and capricious”); *id.* at 1030 (Blackmun, J., dissenting) (arguing that telling the jury about the “rarely exercised power of commutation” has “no greater justification than an instruction to the jury that if the scales are evenly balanced, you should remember that more murders have been committed by people whose names begin with the initial ‘S’ than with any other letter”).
There is ample evidence gathered that capital jurors do indeed impose death based on arbitrary factors in individual cases. The Capital Jury Project, a National Science Foundation-funded multi-state research project, began the unprecedented endeavor of interviewing jurors who served in capital cases. Beginning in 1990, researchers from different disciplines interviewed 1115 jurors in 340 capital trials in 14 different states. Interviewers found that jurors had profound misunderstandings of the law that tilted the balance toward death, including erroneous beliefs that: the law required them to impose death upon finding an aggravating circumstance; aggravating circumstances need not be proven beyond a reasonable doubt; mitigating circumstances did have to be proven beyond a reasonable doubt; and jurors had to agree unanimously on a mitigating circumstance to vote against death. Interviewers also found that jurors exhibited racial bias in imposing death, or chose death simply to avoid a hung jury. They also chose death because they succumbed to pressure.


103 See Marla Sandys, Cross-Overs—Capital Jurors Who Change Their Minds About the Punishment: A Litmus Test for Sentencing Guidelines, 70 Ind. L.J. 1183, 1221 (1995) (stating that the primary concerns of life-to-death cross-overs are “the desire to avoid a hung jury and
from other jurors, including statements by other jurors reducing a sense of responsibility,\textsuperscript{104} because they vastly underestimated the time the defendant would serve if given life,\textsuperscript{105} or because they wrongly believed life without parole was not really life without parole.\textsuperscript{106}

Thus far, we have simply wrung our hands at these findings. After all, this arbitrariness is inevitable in a system designed only to reduce the risk of arbitrariness. However, because such decision making violates the Eighth Amendment command of \textit{Furman}, and the result is to send a man wrongfully to his death, the system has an obligation to address these findings in a legal venue.

\textit{the fear of the defendant’s early release from prison”). In another study, the author related the following exchange between the interviewer and the juror regarding juror’s decision to change his vote from life to death:}

\begin{quote}
Q: You changed?
A: Uh-huh.
Q: What caused you to change?
A: We just quit. It was either that or we deadlocked.
Q: And you didn’t want to be deadlocked.
A: That’s something else that bothered me. Somebody made a comment, I thought of it later. “Well, after all this time and money, how could we vote and have a hung jury?” And that bothered me because I thought when you are voting on somebody’s life, you have to be able to live with the decision—what makes the difference.
\end{quote}


\textsuperscript{104} See Bowers et. al, \textit{supra} note 102, at 252–55, 258 (giving narrative accounts of jurors’ techniques to intimidate pro-lifers, including “isolating a holdout juror, claiming that the law requires a death sentence for the defendant’s crime, asserting the holdout will not be responsible for the defendant’s execution, . . . telling the holdout that the death sentence will be imposed even if he or she does not vote for it[,]” and even coaching the holdout as to how to respond at polling); Hoffman, \textit{supra} note 103, at 1156 (finding that jurors use ways to overcome or avoid their sense of personal moral responsibility, and techniques used against holdouts included telling them their vote was only a recommendation, the law told them what to do, and that they had said in voir dire they could sentence someone to death). Of course, a death sentence founded on a juror’s shifting of responsibility is un-constitutional. See \textit{Caldwell}, 472 U.S. at 332, 340.

\textsuperscript{105} See William J. Bowers & Benjamin D. Steiner, \textit{Death by Default: An Empirical Demonstration of False and Forced Choices in Capital Sentencing}, 77 Tex. L. Rev. 605, 634–38, 671 (1999) (showing that in states that permit parole, capital jurors vastly underestimate the time that first-degree murderers not given death will stay in prison, and that death votes correspond with how much time the juror believes the defendant will serve).

B. The Need for a Record

In order to expose arbitrary decision making, we need a record. In administrative law, there is a built-in procedure for reviewing an agency’s decision for “arbitrary and capricious” decisions: the decisionmaker must make a complete and thorough record of the reasons for the decision.\(^{107}\) Although the review for arbitrariness is narrow, the requirement of a record is broad. In a seminal administrative law decision, the Court stressed that if the “bare record” does not “disclose the factors that were considered or the [decisionmaker’s] construction of the evidence it may be necessary for the District Court to require some explanation” and to “require the administrative officials who participated in the decision to give testimony explaining their action.”\(^{108}\)

Peppered throughout the Court’s death penalty cases is evidence of the Justices grappling with the absence of a record. A pattern emerges where the majority and dissent guess at the impact on the capital jury of some improper factor.\(^{109}\) In *California v. Brown*, the majority and the dissent disputed whether the “reasonable juror” would have discounted mitigating circumstances in the face of an instruction ordering the jury not to make a decision based on “mere sympathy.”\(^{110}\) In *Caldwell* and in *Mills*, as previously discussed, the majority and the dissent engaged in similar debates. In delivering the majority opinion in *Mills*, Justice Blackmun emphasized, “There is, of course, no extrinsic evidence of what the jury in this case actually thought. We have before us only the verdict form and the judge’s instructions.”\(^{111}\)

\(^{107}\) *See, e.g.*, *Motor Vehicle Mfrs.*, 463 U.S. at 43–44 (finding it “relevant that Congress required a record of the rulemaking proceedings to be compiled and submitted to a reviewing court”).


\(^{109}\) *See* Samuel R. Gross, *Race, Peremptories, and Capital Jury Deliberations*, 5 U. Pa. J. Const. L. 283, 285–86 (2001) (“It is a central theme of many appellate decisions on jury decision making: We don’t want to know . . . Once we decide that we can’t know what juries do, we may safely ‘presume’ that whatever it is, it’s okay. But what about death penalty cases? . . . Should we not be more careful to make sure that in these difficult and extremely important cases jurors understand the law correctly and apply it fairly?“); see also Weisberg, *supra* note 16, at 346 (“[W]e can only find error harmless when we can reconstruct the jury’s likely reasoning to determine whether changing one variable in the trial would or might have changed the result. If a penalty decision is a subjective one involving ‘myriad’ and ‘countless’ factors, however, an appellate court can never be sure that adding or subtracting any one factor could not possibly have affected the result.”)

\(^{110}\) 479 U.S. 538, 541–43 (1987); *id.* at 548 (Brennan, J., dissenting).

\(^{111}\) 486 U.S. at 381; see also *Boyd v. California*, 494 U.S. 370, 381–82 (1990) (holding there was no “reasonable likelihood” that jurors misunderstood the judge’s instruction as precluding the consideration of mitigating evidence of the defendant’s background and
Justice Marshall criticized appellate review based on guesswork in his dissent in *Barclay v. Florida*. After a majority of the Court decided that a state trial court’s consideration of an improper nonstatutory aggravating factor did not result in the “wholly arbitrary” imposition of death, Justice Marshall lamented that “[p]rotecting against the arbitrary imposition of the death penalty ‘must not become simply a guessing game played by a reviewing court,’” and that “[i]f appellate review is to be meaningful, it must fulfill its basic historic function of correcting error in the trial court proceedings. A review for correctness reinforces the authority and acceptability of the trial court’s decision and controls the adverse effects of any personal shortcomings in the initial decisionmaker.” It was easy for Justice Marshall thereafter to complete his argument that meaningful appellate review would have made a difference in *Barclay*, however, because the trial judge who made the decision to impose death, against the recommendation of the jury, placed his reasons on the record.

---

113 Id. at 950–51.
115 Id. at 988–89.
116 Because of this record, Justice Marshall was able to expose in dissent that the trial judge had a pattern of making the same “lawless” determination in other cases—ignoring the jury’s advisory sentence of death, recounting his experiences during WWII and reciting “boilerplate language to the effect that he was not easily shocked but that the offense involved shocked him.” Id. at 980–81.
Of course, there is such a record in noncapital sentencing. Although overturning a trial judge’s sentence in a noncapital case is rare because a judge has complete discretion to sentence within a statutory range, there are a few examples where a higher court has reversed a sentence imposed by a trial judge. Then-Justice Rehnquist reviewed two of those decisions in Zant v. Stephens.\footnote{117} United States v. Tucker\footnote{118} and Townsend v. Burke\footnote{119} were two noncapital sentencing decisions vacated for resentencing. In both cases, the defendant was able to show an error of constitutional magnitude in the sentencing because the trial judge had placed his reasons for the sentence on the record. In Tucker, two uncounseled convictions were introduced against the defendant in sentencing.\footnote{120} Justice Rehnquist noted that the Court there observed that “the sentencing judge gave ‘explicit’ and ‘specific’ attention to these convictions” and, had the judge known the true character of convictions, he would have had a “‘dramatically different’” picture of the defendant.\footnote{121} Justice Rehnquist then noted that in Townsend, “an uncounseled defendant was sentenced following a proceeding in which the trial judge explicitly and repeatedly relied upon the incorrect assumption that the defendant had been convicted of several crimes.”\footnote{122}

Ironically, Justice Rehnquist cited those two cases as support for his argument that every capital defendant has an opportunity to prove a claim that the sentencer would have acted differently had the error not occurred, and that the defendant in Zant failed to meet that burden.\footnote{123} Yet, the defendant had no such opportunity in Zant because he had no access to the jurors’ reasons for imposing death.\footnote{124} Thus, in Zant, when the jury imposing death was found to have relied on an

\begin{footnotes}
\footnote{117} 462 U.S. at 902–03 (Rehnquist, J., concurring).
\footnote{118} 404 U.S. 443 (1972).
\footnote{119} 334 U.S. 736 (1948).
\footnote{120} 404 U.S. at 447.
\footnote{121} Zant, 462 U.S. at 902 (Rehnquist, J., concurring) (quoting Tucker, 404 U.S. at 444, 447–48); see Tucker, 404 U.S. at 447 (“The record in the present case makes evident that the sentencing judge gave specific consideration to the respondent’s previous convictions . . . .”).
\footnote{122} Zant, 462 U.S. at 903 (Rehnquist, J., concurring); see Townsend, 334 U.S. at 740 (“We are not at liberty to assume that items given such emphasis by the sentencing court, did not influence the sentence . . . .”).
\footnote{123} 462 U.S. at 903–04 (Rehnquist, J., concurring).
\footnote{124} See id. at 864–68. Likewise, in McCleskey v. Kemp, there was simply no ability to make the proof of racism the Court sought. The Court acknowledged as much: “Controlling considerations of public policy dictate that jurors cannot be called to testify to the motives and influences that led to their verdict.” 481 U.S. at 296 (internal quotations and citations omitted).
\end{footnotes}
aggravating factor subsequently deemed invalid, a majority of the Court upheld the death sentence based on guesswork that the aggravating circumstance had “an inconsequential impact on the jury’s decision regarding the death penalty.”\textsuperscript{125} In dissent, Justice Marshall disputed that assumption: “There is no way of knowing whether the jury would have sentenced respondent to death if its attention had not been drawn to the unconstitutional statutory factor.”\textsuperscript{126}

If noncapital defendants can challenge their sentences through access to the sentencer’s rationale, capital defendants should have this ability, for even more important reasons. Although it is unlikely that a review of a trial judge’s sentence in a noncapital process will bear fruit, the capital sentencing process is different. First, any room for arbitrariness or error in the sentencing scheme means the difference between life and death and not the difference between ten years and twenty years. In imposing the death penalty, there should be as little margin for error as humanly possible.\textsuperscript{127} Second, nonarbitrary imposition is constitutionally required in capital cases by the Eighth Amendment.

The reason we have no record of jurors’ reasons for imposing death is simple, and yet quite foreboding. To suggest scrutiny of a jury verdict is to suggest something quite un-American\textsuperscript{128} and mutinous. As a nation, our rhetoric in committing to jury secrecy is not unlike our commitment to every word in the Pledge of Allegiance. The defenses to both are steeped in history, tradition, and mysticism or spiri-

\textsuperscript{125} 462 U.S. at 889 (quoting Zant v. Stephens, 297 S.E.2d 1, 4 (Ga. 1982)).
\textsuperscript{126} Id. at 905 (Marshall, J., dissenting).
\textsuperscript{127} See Woodson v. North Carolina, 428 U.S. 280, 305 (1976) (“Death, in its finality, differs more from life imprisonment than a 100-year prison term differs from one of only a year or two. . . . [T]here is a corresponding difference in the need for reliability in the determination that death is the appropriate punishment in a specific case.”); see also Eddings, 455 U.S. at 118 (O’Connor, J., concurring) (stating that the Court’s obligation has been “to ensure that the prisoner sentenced to be executed is afforded process that will guarantee, as much as is humanly possible, that the sentence was not imposed out of whim, passion, prejudice or mistake”).
\textsuperscript{128} Indeed, in 1955, upon hearing that Professor Harry Kalven, Jr., was planning to videotape actual jury deliberations as part of a University of Chicago study, the Internal Security Subcommittee of the Committee on the Judiciary held hearings to consider the issue of jury “bugging,” and questioned members of the research team about their associations with communism. See Nancy S. Marder, Deliberations and Disclosures: A Study of Post-Verdict Interviews of Jurors, 82 IOWA L. REV. 465, 530 & n.344 (1997) (describing incident). As a result of the hearings, Congress passed a statute precluding tape recording of jury deliberations. See 18 U.S.C. § 1508 (2000).
tuality. A jury without secrecy is like a nation without God. Each provides a protective layer, or at least a veneer, to our sense of well being. There is a fear that if we scratch beneath the surface, a profound meaninglessness might reveal itself.

Yet, surprisingly perhaps, the rhetoric of jury secrecy is much stronger than the reality. The commitment to jury secrecy post-verdict is uneven and many of the reasons supporting it are surprisingly weak. Nonetheless, to suggest a wholesale review of jury secrecy is impractical and unwarranted. On the other hand, to consider here, for the first time, whether the Eighth Amendment commands easing the demands of jury secrecy in death penalty sentencing proceedings is a realistic and worthwhile endeavor.

IV. OVERCOMING THE BARRIERS AGAINST REVIEWING JURY SENTENCING

The origins of jury secrecy are murky, at best, but it has been an assumed concomitant of the jury trial since its inception. Today’s practice of jury secrecy takes three basic forms. First, the jury deliberates in secret, with no recording or transcription device in the room. Second, evidentiary rules prohibit jurors from testifying

129 See William Holdsworth, A History of English Law 317 (A.L. Goodhart & H.G. Hanbury, eds., 7th ed. 1956) (“The jury is regarded as a formal test to which the parties have submitted. The judgment follows, as under the old system, the result of that test. But to ask in what manner one of the old tests worked, to lay down rules for its working, would have been almost impious; for are not the judgments of God past finding out? The record tells us that when the jury was first introduced the method by which it arrived at its verdict inherited the inscrutability of the judgments of God.”); Mark S. Brodin, Accuracy, Efficiency, and Accountability in the Litigation Process—The Case for the Fact Verdict, 59 U. Cin. L. Rev. 15, 44 (1990) (“The general verdict is as inscrutable and essentially mysterious as the judgment which issued from the ancient oracle of Delphi.”) (quoting Edson R. Sunderland, Verdicts, General and Special, 29 Yale L.J. 253, 258 (1920)).

130 As Justice Cardozo lamented, it was a practice with “its origin in inveterate but vague tradition, and where no attempt has been made either in treatise or in decisions to chart its limits with precision.” Clark v. United States, 289 U.S. 1, 13 (1933).

131 See Abraham S. Goldstein, Jury Secrecy and the Media: The Problem of Postverdict Interviews, 1993 U. Ill. L. Rev. 295, 295 (“For most of the petit jury’s long history, and certainly since the 19th century, the secrecy of jury deliberations has been taken for granted. Like the outcome in trial by ordeal or trial by combat, the jury verdict has been regarded as divinely inspired.”).

about anything that occurred in the jury room with the exception of the intrusion of an extraneous influence.¹³³ Third, local rules in many jurisdictions place restrictions on litigants interviewing jurors post-verdict about their deliberations.¹³⁴ While the latter two forms are far from consistent in their enforcement of jury secrecy, the end effect of all three is to protect against the doomsday prediction that to reveal jury deliberations would be to end the jury system as we know it.¹³⁵

Juries are inefficient decisionmakers, but we value them for a host of reasons related to their messy humanity.¹³⁶ Juries interpose the community between the State and the accused citizen.¹³⁷ Juries are desirable because they bring twelve different, and fresh, views to the table.¹³⁸ With the positive aspects of humanity, however, also come all


¹³³ See, e.g., Fed. R. Evid. 606(b).


¹³⁵ See Tanner v. United States, 483 U.S. 107, 120 (1987) (“There is little doubt that postverdict investigation into juror misconduct would in some instances lead to the invalidation of verdicts reached after irresponsible or improper juror behavior. It is not at all clear, however, that the jury system could survive such efforts to perfect it.”); Goldstein, supra note 131, at 314 (quoting William R. Cornish, The Jury 258 (1968)) (“We proceed at our peril, therefore, both constitutionally and functionally, when we challenge one of the jury’s core characteristics. The inscrutability of the jury verdict, and the secrecy through which it is maintained, ‘is surely [such] a characteristic . . . [and] is bound to last as long as the jury system itself. Once the inscrutability principle has gone, the time has come to set up another kind of tribunal.’”) (alteration in original); see also John H. Wigmore, A Program for the Trial of Jury Trial, 12 J. Am. Jud. Soc’y 166, 170 (1929) (“The jury, in the privacy of its retirement, adjusts the general rule of law to the justice of the particular case. . . . The jury, and the secrecy of the jury room, are the indispensable elements in popular justice.”).

¹³⁶ See Peter N. Thompson, Challenge to the Decisionmaking Process—Federal Rule of Evidence 606(b) and the Constitutional Right to a Fair Trial, 38 Sw. L.J. 1187, 1219–20 (1985) (“Juries bring to the jury room a full range of human characteristics and weaknesses that preclude a totally sterile, objective, rational analysis of the evidence and instructions.”).

¹³⁷ See Duncan v. Louisiana, 391 U.S. 145, 156 (1968) (“Providing an accused with the right to be tried by a jury of his peers gave him an inestimable safeguard against the corrupt or overzealous prosecutor and against the compliant, biased, or eccentric judge. If the defendant preferred the common-sense judgment of a jury to the more tutored but perhaps less sympathetic reaction of the single judge, he was to have it.”).

¹³⁸ See Hans Zeisel, . . . And Then There Were None: The Diminution of the Federal Jury, 38 U. Chi. L. Rev. 710, 715 (1971) (“The jury system is predicated on the insight that people see and evaluate things differently. It is one function of the jury to bring these divergent perceptions and evaluations to the trial process.”).
forms of prejudice, preconception, and nearsightedness.\textsuperscript{139} Supporters of jury secrecy say we must take the good with the bad.\textsuperscript{140} To open up jury deliberations for scrutiny, they say, would threaten finality, undermine the role of the juror as the primary fact finder, encourage jury tampering and harassment, suppress freedom of discourse among jurors, and undermine public confidence in the jury system.

Although these arguments have varying levels of merit in the context of any criminal jury trial—indeed, some appear rather “thin”\textsuperscript{141}—they are particularly difficult to defend in a capital case. Unlike any constitutional safeguard in a noncapital trial, the Eighth Amendment commands focus on the substantive results of the process in the penalty phase of a capital trial.\textsuperscript{142} Because, as a constitutional matter, a jury may not impose a death sentence arbitrarily or capriciously, a defendant must be able to meaningfully challenge the jury’s reasoning, and can do so only with access to that reasoning.\textsuperscript{143}

\textsuperscript{139} The risk of juries injecting their own views or beliefs has long been recognized. As Oliver Wendell Holmes once observed, the jury’s strength in keeping “the administration of the law in accord with the wishes and feelings of the community” is also “precisely one of their gravest defects from the point of view of their theoretical function: that they will introduce into their verdict a certain amount—a very large amount, so far as I have observed—of popular prejudice.” Oliver Wendell Holmes, \textit{Law in Science and Science in Law}, in \textit{Collected Legal Papers} 210, 237–38 (1920).

\textsuperscript{140} See, e.g., United States v. McKinney, 429 F.2d 1019, 1022–23 (5th Cir. 1970) (“We cannot expunge from jury deliberations the subjective opinions of jurors, their attitudinal expositions, or their philosophies. These involve the very human elements that constitute one of the strengths of our jury system . . . .”); Patrick E. Higginbotham, \textit{Juries and the Death Penalty}, 41 Case W. Res. L. Rev. 1047, 1056 (1991) (“There is a point at which we must either accept the irreducible core of discretion inherent in the function of juries or confess that we do not want juries making the decision at all.”); Sandra D. Jordan, \textit{The Criminal Trial Jury: Erosion of Jury Power}, 5 \textit{Howard Scroll: The Soc. Just. L. Rev.} 1, 56 (2002) (“Individual bias is a part of human nature and cannot ever be eliminated through voir dire, blue ribbon juries, or professional juries. Human nature is just that—human.”); Vanessa L. Bellino, Note, \textit{Is the Power to be Lenient Also the Power to Discriminate? An Analysis of Justice Blackmun’s Evolving Perspective on Jury Discretion in Capital Sentencing}, 5 \textit{Temp. Pol. & Civ. Rts. L. Rev.} 75, 84 (1995) (“[W]e must either accept the jury for what it is—a collection of human beings vulnerable to human biases and prejudices—or eliminate its role in the capital sentencing process.”).

\textsuperscript{141} See Alschuler, \textit{supra} note 69, at 227 (“The justifications offered for the rule against the impeachment of jury verdicts by jurors seem thin, and one may wonder whether this rule has served other goals that courts have been reluctant to avow.”).

\textsuperscript{142} In a noncapital trial, the due process and Sixth Amendment rights to a fair trial emphasize fair procedure and not results. Litigants have had little success using the Sixth Amendment to challenge the evidentiary prohibitions on revealing jury deliberations. See, \textit{e.g.}, \textit{Tanner}, 483 U.S. at 126–27 (holding that denying juror testimony of drug and alcohol use during trial does not violate the Sixth Amendment).

\textsuperscript{143} See Higginbotham, \textit{supra} note 134, at 1049 (“[R]esponding to perceptions that the death penalty is imposed in a capricious and irrational manner requires an inquiry such as
This Part first discusses the current status of procedures and practices surrounding jury secrecy, showing that in actuality the legal system’s commitment to secrecy post-trial is not inviolate. Next, this Part shows that the constitutional concerns in a capital trial substantially outweigh the modern rationales supporting secrecy rules.

A. The Current Status of Jury Secrecy

The legal system’s commitment to secrecy of jury deliberations is quite divided: although the jury deliberates in absolute secrecy, the treatment of that secrecy post-trial is riddled with exceptions, inconsistencies, and confusion. Commitment to jury secrecy post-trial can be measured by courts’ adherence to Federal Rule of Evidence 606(b), or its state counterparts, which prohibits jurors from impeaching their own verdicts through testimony. Application of the rule demonstrates that courts are torn over its inhibition of the truth-finding function of a criminal trial.

144 Federal Rule of Evidence 606(b) states, in part:

[A] juror may not testify as to any matter or statement occurring during the course of the jury’s deliberations or to the effect of anything upon that or any other juror’s mind or emotions as influencing the juror to assent to or dissent from the verdict or indictment or concerning the juror’s mental processes in connection therewith, except that a juror may testify on the question whether extraneous prejudicial information was improperly brought to the jury’s attention or whether any outside influence was improperly brought to bear upon any juror.

145 Various judges have expressed their views on the illogic of the rule over time. Shortly after the rule was imported into the common law, Tennessee’s Judge Whyte, in Crawford v. State, decried its existence:

But it is said in the argument, the receiving [sic] the affidavit of jurors is against public policy; it would expose them to the [sic] being tampered with, the effect of which would be numerous applications to set aside verdicts. The like objection applies to every witness—the possibility of being practiced upon. . . . A verdict under such circumstances is to be approached with great caution and great circumspection; but it is not altogether intangible, and beyond the reach of the redressing power of the court; if it were, I for one would think it a defect in the administration of the justice of the country, and a defect in the policy of the law.

10 Tenn. (2 Verg.) 60, 67, 69 (1821). Judge Learned Hand advised courts to avoid the rule that jurors are incompetent to testify to impeach their verdict, and to look at the facts. Jorgensen v. York Ice Mach. Co., 160 F.2d 432, 435 (2d Cir. 1947). Chief Justice, then Judge, Burger agreed:
Rule 606(b) is not an absolute prohibition on juror testimony about occurrences in the jury room. It prohibits juror testimony on the mental processes involved in deliberations, but it does admit juror testimony on extraneous influences on the jury verdict.\textsuperscript{146} Testimony that is typically precluded under the Rule includes evidence of juror misunderstanding or intentional disregard of court instructions, a verdict achieved through compromise, consideration of the defendant’s failure to testify or of inadmissible evidence, misgivings about the verdict, agreement on a time limit for decision, inability to hear or comprehend the trial, coercion or harassment of fellow jurors, lack of intention to unqualifiedly vote guilty, basing the verdict on secret beliefs or prejudices unrelated to the law or facts, mistake in returning verdicts, and improper inferences from a codefendant’s guilty plea.\textsuperscript{147}

The list of excluded testimony is contrary to both the truth-finding and fairness goals of a criminal trial. If one reads through the cases rejecting juror testimony under Rule 606(b), uncorrected miscarriages of justice leap off of the page: jurors deciding to convict based on visions,\textsuperscript{148} due to intense pressure from other ju-

\textsuperscript{146} The reason for the distinction was that, although allowing testimony as to outside influences was capable of proof, \[t\]o allow a juror to make affidavit against the conclusiveness of the verdict by reason of and as to the effect and influence of any of these matters upon his mind, which in their very nature are, though untrue, incapable of disproof, would be practically to open the jury room to the appliances of parties and their attorneys, and, of course, thereby to unsettle verdicts and destroy their sanctity and conclusiveness.


\textsuperscript{148} See United States v. Dioguardi, 492 F.2d 70, 75 (2d Cir. 1974) (holding inadmissible evidence that a juror wrote the defendant that she had “eyes and ears that . . . see things before it [sic] happen” and claiming that “a curse was put on them some years ago”); Hutchinson v. Laughlin, 102 N.E.2d 875, 879 (Ohio Ct. App. 1951) (holding inadmissible evidence of an astrological investigation by the jury foreperson); State v. DeMille, 756 F.2d...
rors, due to time constraints, or simply based on mistake. What is so striking about this practice is that it comes at the tail end of a trial stacked with procedures designed to ensure a reliable outcome—“illustrating a central dynamic of American criminal justice: Millions for procedure but not one dime for outcome.” The decision to preclude this testimony “reflects a policy judgment that the interest in accurate and rational decision making is sub-

81, 83–84 (Utah 1988) (excluding evidence that a juror claimed she had received a divine sign of the defendant’s guilt).

In United States v. Roach, the Eighth Circuit Court of Appeals held inadmissible a juror affidavit attesting to juror pressure. The court noted that the affidavit stated that the juror “had been unwilling to convict but that the other jurors had pressured her into changing her vote” and that “one juror told her the judge would incarcerate her if she failed to do her civic duty and vote to convict.” 164 F.3d 403, 413 (8th Cir. 1998) (paraphrasing affidavit). In addition, the court stated the juror’s affidavit indicated that “there were racial overtones in the jury room,” she was “one of two Native American jurors, and for a time she was the only holdout against convicting the three Native American defendants. She said other jurors made references to her race and one said ‘[i]t was ten white people versus one Indian.” Id. The court further noted that the juror’s affidavit indicated she was diabetic and that “other jurors told her that she could get something to eat with them after a verdict was returned.” Id.; see also United States v. Brito, 136 F.3d 397, 414 (5th Cir. 1998) (holding inadmissible a juror affidavit stating that the juror experienced “internal coercion” due to harassment by other jurors); United States v. Tallman, 952 F.2d 164, 166 (8th Cir. 1991) (holding inadmissible a juror’s claims of “harassment and insults” from other jurors); United States v. Norton, 867 F.2d 1354, 1366 (11th Cir. 1989) (holding inadmissible a juror’s claim of experiencing “duress” during deliberations due to alleged harassment or intimidation by other jurors); United States v. Barber, 668 F.2d 778, 786 (4th Cir. 1982) (holding there was no basis to impeach a verdict where a juror claimed the foreman “scared [her] to death”) (alteration in original); Anderson v. Miller, 206 F. Supp. 2d 352, 360–62 (E.D.N.Y. 2002) (finding inadmissible testimony from two jurors, who cried and were almost inaudible during polling, and who said they had been coerced and intimidated by fellow jurors); Oxtoby v. McGowan, 447 A.2d 860, 869–70 (Md. Ct. Spec. App. 1982) (holding inadmissible a juror affidavit claiming coercion); State v. Franklin, 534 S.E.2d 716, 719 (S.C. Ct. App. 2000) (holding inadmissible a juror affidavit claiming that the juror had been coerced to vote guilty).

See United States v. Graveley, 840 F.2d 1156, 1159 (4th Cir. 1988) (holding inadmissible evidence that jurors, who had requested an opportunity to comment on their verdicts but were denied, told press that they thought defendant innocent but rendered the guilty verdicts because of extreme time pressure).

See United States v. Weiner, 578 F.2d 757, 764 (9th Cir. 1978) (holding inadmissible a juror’s statement to the judge thirty minutes after the verdict that she had only voted guilty with reservations); Aguilar v. State, 242 S.E.2d 620, 623 (Ga. 1978) (excluding affidavits of three jurors stating they thought the defendant should have been convicted of manslaughter even though they voted to convict him of murder).

Alschuler, supra note 69, at 226; see also Ruprecht, supra note 132, at 243 (“There is something comically inconsistent about the scrupulous observation of proper procedure and formal proof during the trial, and the final submission of the cause to a deliberative body that operates free of all rational constraint with a request that it return a verdict unjustified by reasons.”).
servient to the interest of finality” and “represents a commitment to democratic decision making and not necessarily a commitment to finding the truth.”

In a noncapital case, the decision to favor the finality of the jury’s decision in order to preserve this democratic institution has an inevitable appeal. Therefore, in 1915, in *McDonald v. Pless*, the Court set the tone for the rule by declaring that, although the jurors in the case would have testified that they “adopted an arbitrary and unjust method in arriving at their verdict,” forgoing this proof was “the lesser of two evils.” In a capital case, however, those values are reversed. Fifty-seven years after *McDonald*, the Court in *Furman v. Georgia* identified arbitrary and unjust methods for arriving at death as a primary evil in capital trials. In capital cases applying Rule 606(b), lower courts have rejected testimony that, if allowed, would have shown that a juror arbitrarily imposed death, including testimony that jurors voted for death because of speculation that the defendant would be released in a few years if given life and not executed if given death, or that the governor would commute a life sentence, or because of a feeling of reduced responsibility because it was a retrial, or due to coercion or mistake.

---

153 Thompson, *supra* note 136, at 1188.
154 238 U.S. 264, 267 (1915).
155 See generally 408 U.S. 238 (1972) (per curiam).
156 See *McDowell v. Calderon*, 107 F.3d 1351, 1367 (9th Cir. 1997), *vacated in part*, 130 F.3d 833 (9th Cir. 1997) (en banc) (excluding affidavit of holdout juror that stated that other jurors had become increasingly angry with her, told her that “without parole” meant the defendant could be paroled within ten to fifteen years, and asserted the defendant would not be executed because no one was being executed in California, and that she had voted for death because she did not want him released and did not believe he would be executed); *Silagy v. Peters*, 905 F.2d 986, 1008–09 (7th Cir. 1990) (holding inadmissible testimony that during the sentencing phase, one juror told others that the defendant would serve no more than five to seven years if they sentenced him to life, and that even if they gave him death, he would never be executed, but would spend no more than seven years in prison).
157 See *Bloom v. Vasquez*, 840 F. Supp. 1362, 1377–78 (C.D. Cal. 1993), *rev’d on other grounds*, 143 F.3d 1267 (9th Cir. 1997), (holding inadmissible evidence that one juror told others that with a life sentence defendant would be eligible for parole in seven years, that several jurors opined he would never be executed, and that two jurors who initially were in favor of life voted for death after a discussion among the jurors regarding the possibility of life being commuted by the governor).
158 See *Fullwood v. Lee*, 290 F.3d 663, 682–83 & n.8 (4th Cir. 2002) (holding that, although a court could admit evidence that the jury became aware from an outside source that defendant had already been sentenced to death by another jury and the sentence had been reversed on a technicality, a statement by a juror regarding the effect of this knowledge was not admissible, where the juror indicated that it “lessen[ed] our sense of responsibility . . . because we felt that twelve other rational people had sentenced Mr. Fullwood to death.”) (alterations in original).
159 See *Gosier v. Welborn*, 175 F.3d 504, 510–11 (7th Cir. 1999) (excluding juror’s claim that he voted for death because he could not convince other jurors to be lenient and appar-
Even if a rule of post-verdict secrecy is defensible in noncapital cases, the courts are not always consistent in enforcing its boundaries. Courts’ manipulation of the basic terms of Rule 606(b) demonstrates the lack of integrity in the Rule. For example, courts have been inconsistent in their determinations of whether something was an internal or an external influence.\(^\text{160}\) The easy manipulation of this feature of the Rule is demonstrated by the widely criticized decision of *Tanner v. United States*, where the Supreme Court held that juror intoxication during trial was not an extraneous influence.\(^\text{161}\) Without considering juror testimony as to the existence, much less the effect, of this impropriety, the Court upheld the defendants’ convictions.\(^\text{162}\)

Even if a court were to find that intoxication were an external influence, however, the Rule precludes juror testimony as to its effect.\(^\text{163}\) Instead of simply allowing the juror to testify to the effect, judges guess.\(^\text{164}\) In *Sassounian v. Roe*, the Ninth Circuit admitted that ignoring a juror’s own testimony that extraneous information caused her to vote to find the defendant eligible for death lent an “Alice in
Wonderland quality’” to the court’s determination of the effect that the extraneous information had upon her.\textsuperscript{165}

Perhaps because of the idiosyncrasies of the Rule in operation, the Supreme Court has simply ignored its implications on occasion. In \textit{Smith v. Phillips}, the Court found that due process required that a hearing be held to resolve a claim of juror bias.\textsuperscript{166} The Court affirmed the trial court’s finding of no bias, based in part on the testimony of the allegedly biased juror that the supposed bias did not affect his deliberations.\textsuperscript{167} Rule 606(b) would have precluded such a hearing, but the Court did not even mention the Rule.\textsuperscript{168} Similarly, in \textit{Rushen v. Spain}, the Court gave short shrift to Rule 606(b) in a footnote\textsuperscript{169} and held there was no constitutional violation due to a juror’s personal knowledge of the crime, based in part on her testimony in a post-trial hearing that it did not affect her impartiality.\textsuperscript{170} In both cases, the Court easily dispensed with the claim by hearing from the juror herself.\textsuperscript{171}

The foregoing discussion of Rule 606(b) demonstrates that commitment to jury secrecy, at least post-trial, is not as monolithic as might have been assumed. The Rule allows inquiry into some areas while forbidding inquiries into others, and courts’ application of the Rule has been less than consistent and logical. The basic form of the Rule, however, will continue to be supported by Congress\textsuperscript{172} because...

\textsuperscript{165} 230 F.3d 1097, 1109 (9th Cir. 2000) (quoting People v. Sassounian, 226 Cal. Rptr. 880, 914 (Ct. App. 1986)).
\textsuperscript{166} 455 U.S. 209, 218 (1982).
\textsuperscript{167} \textit{See id.} at 218–20.
\textsuperscript{168} \textit{See id.} at 222 (O’Connor, J., concurring) (supporting necessity of post-trial hearing to resolve claims of juror bias because “[a] hearing permits counsel to probe the juror’s memory, his reasons for acting as he did, . . . his understanding of the consequences of his actions[,] . . . [and] also permits the trial judge to observe the juror’s demeanor under cross-examination and to evaluate his answers in light of the particular circumstances of the case”).
\textsuperscript{169} 464 U.S. at 121 n.5.
\textsuperscript{170} \textit{Id.} at 120–21.
\textsuperscript{171} Courts have also avoided the internal-external distinction in the face of evidence of racial bias. \textit{See}, e.g., Wright v. United States, 559 F. Supp. 1139, 1151 (E.D.N.Y. 1983) (finding “potential constitutional difficulties in applying Rule 606(b) to all allegations of racial prejudice” and the “better rule . . . is to analyze each such claim on a case by case basis”); Smith v. Brewer, 444 F. Supp. 482, 490 (S.D. Iowa 1978), aff’d 577 F.2d 466 (8th Cir. 1978), (holding that the rule should not “be applied dogmatically” where an offer of proof showed racial bias in the jury room).
\textsuperscript{172} Indeed, in May 2004, the Advisory Committee on the Evidence Rules recommended approval of an amendment to Rule 606(b) which would clarify the narrowness of the scope of the exception for impeachment of the verdict. \textit{See Advisory Committee on
its persistence is tied to its symbolic quality as the guardian of jury secrecy. On the other hand, no one has really questioned its application in capital cases in light of Eighth Amendment commands. The political, judicial, and symbolic costs of lifting such a rule in capital cases is substantially less. Further, although the rationales given in support of jury secrecy may justify its use in noncapital cases, they simply cannot justify its use in capital cases.

B. The Modern Rationale for Jury Secrecy

1. Finality

Finality is a powerful justification for jury secrecy. In *McDonald*, one of the first modern cases endorsing the evidentiary ban on jury testimony impeaching verdicts, the Supreme Court proclaimed the virtue of finality:

> [L]et it once be established that verdicts solemnly made and publicly returned into court can be attacked and set aside on the testimony of those who took part in their publication and all verdicts could be, and many would be, followed by an inquiry in the hope of discovering something which might invalidate the finding.

Of course, verdicts can and are attacked through jury testimony about extraneous influences on the verdict under Rule 606(b). Nonetheless, there is still a blanket preclusion of inquiry into the mental processes of jurors. The concern is very real that verdicts are delicately arrived at and maintained and, given the vagaries and emotions

---


174 This Article does not discuss either the ancient rationale supporting jury secrecy—that the jury was divinely inspired—or the rationale that first supported the importation of a no impeachment rule into common law—that jurors were incompetent to testify to their own moral turpitude. *See Holdsworth, supra* note 129, at 317; Brodin, *supra* note 129, at 44. *See generally* Vaise v. Delaval, 99 Eng. Rep. 944 (K.B. 1785). Neither rationale supports the practice today. *See, e.g.*, Smith v. Cheetham, 3 Cai. R. 57, 59 (N.Y. Sup. Ct. 1805) (repudiating reasoning by Mansfield because juror evidence of misconduct is “the best and highest evidence”).

175 238 U.S. at 267.
of twelve separate human beings, no verdict could withstand the scrutiny.\textsuperscript{176}

In the context of capital cases, however, finality must take a back seat to overriding constitutional concerns.\textsuperscript{177} Finality has a unique meaning in a capital case. Because death itself is final, the Supreme Court has recognized a heightened concern for reliability in the process. Finality in death weighs in favor of more scrutiny, not less.\textsuperscript{178} As Justice Breyer recently said, “I believe we should discount ordinary finality interests in a death case, for those interests are comparative in nature and death-related collateral proceedings, in any event, may stretch on for many years . . . .”\textsuperscript{179}

A protracted appellate process is already in place to ensure the protection of a capital defendant’s constitutional rights.\textsuperscript{180} In a death case, the sentence is not final until the defendant is executed, and the appellate process often continues up until the moment of execution. Disallowing litigation of an Eighth Amendment claim by a capital de-

\textsuperscript{176} See Jorgensen, 160 F.2d at 435 (“[I]t would be impracticable to impose the [requirement] of absolute perfection that no verdict shall stand, unless every juror has been entirely without bias, and has based his vote only upon evidence he has heard in court. It is doubtful whether more than one in a hundred verdicts would stand such a test . . . . [judges] would become Penelopes, forever engaged in unravelling the webs they wove.”); Crump, \textit{supra} note 134, at 534 (“The granting of new trials based on jurors’ mental processes is prohibited because it would be anathema to stability; indeed, the setting aside of jury verdicts on any but the most egregious grounds would cost more in terms of stability and finality than it could possibly gain.”); Thompson, \textit{supra} note 136, at 1187 (“[T]oo close a look at jury deliberations will reveal improprieties in a large number of cases, damaging the finality and public acceptance of jury verdicts.”).

\textsuperscript{177} See Mark Cammack, \textit{The Jurisprudence of Jury Trials: The No Impeachment Rule and the Conditions for Legitimate Legal Decisionmaking}, 64 U. Col. L. Rev. 57, 77–79 (1993) (stating that finality is an important efficiency concern, but it also could justify curtailment of almost any constitutional right).

\textsuperscript{178} See, \textit{e.g.}, Ake v. Oklahoma, 470 U.S. 68, 87 (1985) (Burger, C.J., concurring in the judgment) (“In capital cases the finality of the sentence imposed warrants protections that may or may not be required in other cases.”); California v. Ramos, 463 U.S. 992, 998–99 (1983) (stating that the qualitative difference of death “requires a correspondingly greater degree of scrutiny of the capital sentencing determination”).


fendant, while allowing litigation of other constitutional claims, does not advance finality significantly enough to justify it.

Indeed, having jurors provide reasons for their sentence may actually aid in finality. The appellate process is currently steeped in guesswork about whether a jury did or did not understand its instructions, or whether there was a reasonable likelihood that the jury verdict would have been different.\footnote{For a discussion of this issue, see supra notes 109–16 and accompanying text.} Hearing from the jurors themselves could aid in what is essentially a fact-finding process.

Most importantly, unlike in noncapital cases, defendants in capital cases have an Eighth Amendment right to non-arbitrary imposition of death. The only effective method for realizing this right is to expose arbitrariness in the decision by scrutinizing the jury’s decision-making process. Finality concerns, while important, should not shield from view evidence establishing the arbitrary imposition of death in violation of the Eighth Amendment.

2. Jury as Primary Factfinder

Another rationale for secrecy in deliberations is that it is necessary to preclude encroachment on the jury’s fact-finding function. One primary purpose of the criminal jury is to interpose the community between the criminal defendant and the power of the government.\footnote{See Duncan, 391 U.S. at 156.} The concern is that if a trial court or an appellate court were allowed to make a searching review of the jurors’ reasons for their verdict, it could simply substitute its own ideals for the ideals of the community.\footnote{See, e.g., Chicago, Rock Island & Pac. R.R. Co. v. Speth, 404 F.2d 291, 296 (8th Cir. 1968) (“If a court could attempt to ‘correct’ a verdict by general inquiry of a jury, even though the questioning is well-intended and seemingly innocuous on the surface, the protected cloak of privacy around a jury’s deliberations would be permanently shattered.”); Victor Gold, Juror Competency to Testify that a Verdict was the Product of Racial Bias, 9 St. John’s J. Legal Comment. 125, 136–37 (1993) (“If jurors could be made to testify as to the thought processes that formed the foundation of a verdict, then any exercise of their power inconsistent with the values of the judge could be detected and controlled.”); Thompson, supra note 136, at 1222 (“If the essential role of the jury is to serve as a check on governmental tyranny in the exercise of the laws, then the jurors must be free from governmental scrutiny and possible retaliation.”).}

The foundation of this belief is largely historical. The history of the rise of the jury trial is the story of a power struggle between judges and juries.\footnote{See John H. Langbein, The Criminal Trial Before the Lawyers, 45 U. Chi. L. Rev. 263, 284–300 (1978) (describing history).} Judges initially reached out to change juries’ verdicts of
acquittal. There is an important distinction between this history—which can leave one cold about the ability of judges to bow to the decisions of jurors—and the modern day. Since the nineteenth century, acquittals by juries have been beyond the power of the courts to overturn. Through the double jeopardy clause of the Fifth Amendment, a verdict of not guilty, and, similarly, in the capital penalty context, a decision of life over death, is final.

The dictates of double jeopardy law are consistent with the historical value of the jury as community arbiter. Because of the power struggle over acquittals, the framers believed one of the key values of the jury was its ability to acquit in the face of government overreaching, even if it believed the government had proven a crime. In capital cases, juries exercised this power to nullify in order to temper a mandatory death penalty. The abilities to nullify and show mercy are the unique features of a jury drawn from the defendant’s community. Those features of the jury trial system would remain untouched by a review of a jury’s reasons for imposing death. A jury’s decision to exercise mercy, no matter how arbitrary, is unreviewable.

On the other hand, appellate courts expend a great deal of time and energy reviewing the penalty phase of a capital case for harmless

---

185 See id. at 297–300.
188 See Alschuler & Deiss, supra note 186, at 873–74 (discussing that one of the jury’s historic functions was to nullify, as in the seditious libel cases brought by the English Crown); Nancy Jean King, The American Criminal Jury, 62 Law & Contemp. Probs. 41, 50 (1999) (noting that nullification “provided a shield against British oppression before the Revolution”).
190 See Gregg v. Georgia, 428 U.S. 153, 199 (opinion of Stewart, Powell, and Stevens, JJ.) (writing that nothing in Furman “suggests that the decision to afford an individual defendant mercy violates the Constitution”; see also United States v. Martin Linen Supply Co., 430 U.S. 564, 572 (1977) (holding that the jury has the “overriding responsibility to stand between the accused and a potentially arbitrary or abusive Government that is in command of the criminal sanction” and that “[t]he trial judge is thereby barred from attempting to override or interfere with the jurors’ independent judgment in a manner contrary to the interests of the accused”).
error. Information from the jurors themselves supports the jury’s factfinding function and enhances the accuracy of the court’s decision.191

3. Jury Tampering

Another argument in support of jury secrecy is that allowing jurors to impeach their verdict will encourage harassment of jurors.192 Yet, by far, the most direct and efficient rule for preventing harassment is to prosecute, or impose disciplinary proceedings against, anyone who harasses a juror.193 Juror harassment is already precluded by ethical rules.194 This concern is no different from a concern that lawyers and litigants might browbeat or harass witnesses, urging them to testify in a favorable manner.195 A lawyer or two may well engage in this behavior, but the remedy is to punish those individuals, not to preclude relevant testimony.

Closely linked to this concern is a fear about a more subtle coercion. A juror may already feel uncomfortable about her verdict, particularly a decision to impose death, and a good defense lawyer could convince her to confess to having made the wrong decision due to some external pressure. The argument is that in imposing a death sentence, some jurors will have compromised, because compromise is a necessary ingredient to unanimity, and may have lingering regrets about it that can be unearthed by a sympathetic defense attorney. Precluding jurors from testifying about their motivations prevents those lingering regrets from undermining verdicts.196

191 See Thompson, supra note 136, at 1219 (“[T]he appellate function has become largely a fact-finding function in which the transcript is reviewed and evidence is reassessed. The courts must assess both the appropriate weight and the strength of various inferences to be drawn from the evidence and the possible inferences that the jury drew or might have drawn from the erroneously admitted or excluded evidence in determining what impact the error had or might have had on the deliberations and proceedings.”).

192 See McDonald, 238 U.S. at 267 (“Jurors would be harassed and beset by the defeated party in an effort to secure from them evidence of facts which might establish misconduct sufficient to set aside a verdict.”).

193 Accord Alschuler, supra note 69, at 227.

194 See, e.g., MODEL CODE OF PROF’L RESPONSIBILITY DR 7–108(D) (1980) (“After discharge of the jury from further consideration of a case with which the lawyer was connected, the lawyer shall not ask questions of or make comments to a member of that jury that are calculated merely to harass or embarrass the juror or to influence his actions in future jury service.”).

195 See Thompson, supra note 136, at 1224 (“The process of interviewing jurors after a verdict does not appear to present more of an opportunity for unseemly conduct on behalf of the interviewer than the process of interviewing witnesses prior to the trial.”).

196 See id. at 1220–21 (making this argument).
It is certainly true that convincing twelve citizens to condemn someone to death is a daunting task, and, if achieved, no doubt jurors will develop regrets. Regret over having to carry out this unpleasant task as a citizen does not undermine a death sentence validly imposed. If a juror does utilize this regret in order to fabricate the circumstances of his or her decision, then the fear over post-verdict juror testimony will have been realized. However, a juror may also come forward to testify that the pressure of time, the coercion of his fellow jurors, or the instructions of the judge to break a deadlock caused him to join a decision with which he did not agree. In the latter case, the imposition of death violates Furman’s mandate. To make the decision to impose death because of pressure or coercion is an arbitrary and capricious choice and not a “reasoned moral response.”

Although we may be able to countenance that sort of interplay between twelve human beings in a non-capital context, if it leads to an imposition of death in a capital case, the sentence is unconstitutional. Therefore, the answer to the problem of the fabricating witness lies in the crucible of adversarial testing rather than the suppression of testimony.

4. Free and Open Deliberations

Fostering free and open deliberations is by far the most persuasive reason behind jury secrecy. This has driven the practice from the earliest times. Justice Cardozo stated, “Freedom of debate might be stifled and independence of thought checked if jurors were made to feel that their arguments and ballots were to be freely published to the world.”


198 Clark, 289 U.S. at 13.
theme was echoed by the courts and continues to be a primary rationale today.

This rationale is somewhat undermined by the fact that most jurors are already aware that a fellow juror may speak with the media and reveal jury deliberations. Litigants may be under some constraints by a court order or local rule, but, by and large, they are also free to talk to the jurors about deliberations post-verdict. Therefore, any perceived restraint on free and open discussion in the jury room caused by later revealing to a court jurors’ reasons for imposing death likely would be marginal.

There is still a legitimate concern, however, that if jurors know ahead of time that their sentencing decision in a capital case will be scrutinized by a court, the possibility could have an impact on the deliberations. Yet it would defeat the purpose of enforcing the Eighth

---

199 See, e.g., *McDonald*, 238 U.S. at 267–68 (“[T]he result [of allowing jurors to impeach verdicts] would be to make what was intended to be a private deliberation, the constant subject of public investigation—to the destruction of all frankness and freedom of discussion and conference.”); *Woodward v. Leavitt*, 107 Mass. 453, 460 (1871) (“[I]t is essential to the freedom and independence of their deliberations that their discussions in the jury room should be kept secret and inviolable; and to admit the testimony of jurors to what took place there would create distrust, embarrassment and uncertainty.”); *In re Cochran*, 143 N.E. 212, 213 (N.Y. 1924) (“Public policy requires that [jurors] be given the utmost freedom of debate . . . .”); *In re Nunns*, 176 N.Y.S. 858, 873 (App. Div. 1919) (Putnam, J., dissenting) (“[W]hat a juror says and how he votes is within the seal of secrecy for all time. How could justice be administered through results of free conference, unless jurors understand that their deliberations in the jury room are inviolable, and that the reasons for their verdict cannot be questioned.” (quotations and citations omitted)).

200 One of the reasons is to encourage unanimity. See *M’Kain v. Love*, 20 S.C.L. (1 Hill) 506, 508 (S.C. Ct. App. 1834) (“We know from experience, that in questions admitting of any doubt, the only possible means of arriving at unanimity of opinion amongst many, is by a free interchange of thought, and to deny it to a jury would be to defeat the object of trial by jury.”); *John D. Jackson, Making Juries Accountable*, 50 AM. J. COMP. L. 477, 494 (2002) (stating concern that “jurors would be less willing to reach consensus if they knew that their compromises would be later revealed to the public”).

201 Judges cannot issue orders preventing jurors from voluntarily seeking out the press to talk about their deliberations. See *Marcy Strauss, Juror Journalism*, 12 YALE L. & POL’Y REV. 389, 417–18 (1994); *Thompson*, supra note 136, at 1225 (explaining that jurors are free to talk about their deliberations and courts have been unwilling to enjoin journalists).

202 See *Alschuler, supra* note 69, at 226 (“If the prospect that jurors will recount the conduct of other jurors in these [public] forums does not inhibit frank discussion, discourage jurors from taking unpopular stands, and undermine the public’s confidence in jury verdicts, it seems doubtful that the prospect of disclosure in a court of law would do so.”).

203 Psychologists have determined that “jurors are better able to resist normative pressure when their judgments are made anonymously.” *Saul M. Kassin & Lawrence S. Wrightsman, The American Jury on Trial: Psychological Perspectives* 191 (1988), cited in Kenneth B. Nunn, *When Juries Meet the Press: Rethinking the Jury’s Representative Func-
Amendment proscription against arbitrary imposition of death if the kind of juror comments that most revealed arbitrariness were driven underground due to the prospect of discovery. Therefore, the need to enforce the Eighth Amendment must be balanced against this concern and is best addressed by the method chosen to facilitate the jurors’ disclosure of their reasons for imposing death. For example, placing a video camera in the jury deliberation room is potentially more damaging to juror freedom of expression than granting post-trial interviews with litigants. The selection of a balanced method will be explored further in Part V.204

5. Public Confidence in the Jury

There are two aspects of the argument that jury secrecy is imperative to public confidence in the jury system. First, “the community’s trust in a system that relies on the decisions of lay people would . . . be undermined by a barrage of post-verdict scrutiny of juror conduct.”205 Hence, if the community saw jurors continually coming forward to impeach their own verdicts, the community would despair of the process. This argument depends, first, on the existence of “a barrage.” This is of little concern when confined to capital cases, which are relatively few in number.206 This argument also depends upon an assumption that the public would rather close its eyes than have wrongs righted, an assumption that lies outside the mainstream view of the value of public proceedings.207 In any case, current community

204 For a discussion of methods of recording deliberations and the reasons recording deliberations should be rejected, see infra notes 215–36 and accompanying text.

205 Tanner, 483 U.S. at 121; see Marder, supra note 128, at 498 (arguing that judges and jurors “may have a duty to appear more certain than they feel so that the parties and public will accept the decision”).

206 As Nancy King noted, “[C]apital jury trials are relatively uncommon. Of the 2,000 to 4,000 defendants a year charged with a crime that makes them eligible for the death penalty, only about six to fifteen percent receive a death sentence, an average of about 250 death sentences per year.” King, supra note 188, at 64.

207 “American society looks askance at decision-making that takes place behind closed doors. This is illustrated by the ubiquity of open-meeting laws as well as the prevalence of video cameras in courtrooms and legislatures. The jury deliberating in secrecy on a general verdict represents a dramatic exception.” Brodin, supra note 129, at 105. The public
trust in the results of the capital process may be at a new low because it is now common knowledge that innocent men sit on death row.\textsuperscript{208} Public confidence in the sentence of death may well improve if the jury is required to be accountable for its reasons for imposing death.\textsuperscript{209} It is also widely known to the public that capital cases are reversed, retried, and appealed again because of error. One more ground of error in the form of juror testimony is unlikely to tip the balance against trust in the jury system.

A second component of the argument for jury secrecy is that secrecy is thought necessary to “preserve public confidence in a system which more intimate knowledge might destroy.”\textsuperscript{210} The underlying reason, then, is “the widespread belief that jury deliberations may not live up to an ideal of enlightened exchange of views and sifting of evidence, and that the jury as an institution might not survive close scrutiny of its deliberative process.”\textsuperscript{211} This may well be true, but it is by no means a known quantity. Exposing to appellate scrutiny jurors’ rationales for imposing death will reveal mistakes, bias, and caprice, but it will also reveal thoughtful and serious deliberations. The very concern that we will see what we fear—arbitrariness in imposing


\textsuperscript{209} See Jackson, supra note 200, at 486 ("[A]ccountability can help to legitimate the position of the decision makers, as increasingly in liberal democracies decision making must be transparent if there is to be any public confidence in it. . . . [A]ccountability can enhance respect for the individuals affected by the decisions made as it guarantees them some scrutiny over these decisions.").

\textsuperscript{210} GLANVILLE WILLIAMS, THE PROOF OF GUILT: A STUDY OF THE ENGLISH CRIMINAL TRIAL 268 (3d ed. 1963), quoted in Cammack, supra note 177, at 78.

death—counsels that we open our eyes to the Eighth Amendment violation.\footnote{212 See Brodin, supra note 129, at 41 (“[C]ontroversial questions that are difficult to resolve in the open political arena are shunted into the black box where they can be handled discreetly, beyond close scrutiny.”); Weisberg, supra note 16, at 395 (“The development of the formal model [of the penalty trial], at least in the long run, suppresses more than it answers the moral and political questions that ought to be addressed before we execute people.”).}

In sum, the Eighth Amendment mandate against arbitrary imposition of death is not outweighed by the public policy rationales put forth to support jury secrecy. The Supreme Court has not declared jury secrecy to be of constitutional dimension,\footnote{213 See, e.g., Goldstein, supra note 131, at 297.} and it is unlikely to achieve that status.\footnote{214 The Court has decided that the Sixth Amendment right to a jury trial does not include either a twelve-member jury, Williams v. Florida, 399 U.S. 78, 86 (1970), or a unanimous jury, Apodaca v. Oregon, 406 U.S. 404, 406 (1972), even though both were fixed features of the jury system at the time of the Amendment’s adoption. The operative question is whether the incident is part of the “essential feature” that lies in the “interposition between the accused and his accuser of the commonsense judgment of a group of laymen, and in the community participation and shared responsibility that results from that group’s determination of guilt or innocence.” Williams, 399 U.S. at 100. Lack of complete secrecy in deliberations does not harm that function. As Akhil Reed Amar has posited, “inscrutability and muteness are not the essence of juries.” Akhil Reed Amar, Reinventing Juries: Ten Suggested Reforms, 28 U.C. Davis L. Rev. 1169, 1187 (1995).} The collection of age-old practices protecting the jury’s mental processes from intrusion inhibits the search for justice in capital cases. Shedding those practices in capital cases would cost little and greatly benefit the criminal justice system, capital defendants, and the public. Having shaken off the vestments of jury secrecy, then, the more difficult question is how to implement a meaningful and effective review of the jury’s decision to impose death.

V. METHODS OF ENFORCING THE EIGHTH AMENDMENT RIGHT

There are two distinct methods for gaining information from the jurors about their decision making. The two methods are recording jury deliberations or questioning jurors after the verdict. Although recording obviously occurs at the point of deliberations, questioning can occur at two points—either before the jurors are dismissed, through a post-verdict voir dire or questionnaire, or at any point after the jury has been dismissed, through interviews with the parties. This Article concludes that recording deliberations is wholly inadequate to the task, and that the second method of jury questioning is most likely to reveal more reliable information after the jury has been dismissed.
Although there are some questions of reliability inherent in any interviews with jurors, shedding light on juror arbitrariness is far superior to allowing it to continue in the dark.\textsuperscript{215}

\textbf{A. Recording Jury Deliberations}

There are two basic problems with recording jury deliberations as a method for uncovering arbitrary decision making in capital cases. First, recorded deliberations inform us what was said, but do not inform us what was \textit{not} said. While the recording may produce some useful evidence of bias, pressure, or legal misconceptions, it is likely that some, if not most, prejudice, error, and caprice will be unspoken. In addition, some jurors may not speak because they are timid or silenced.\textsuperscript{216} Others may not feel free to state their reasons out loud because they are in the minority or are feeling pressured by other jurors. A verdict, while collective, is also individual. Although the group may choose to impose death, individuals may not agree on the reasons. A juror may have imposed death because she felt pressure to do so or was mistaken about the law, neither of which may be reflected in a recording. Meaningful appellate review of decisions to impose death requires knowledge of the rationale of each and every juror. Because the decision must be unanimous, a single juror’s arbitrary imposition of death requires a reversal of the sentence.\textsuperscript{217}

Second, if there is one place in the process where secrecy makes some sense, it is during the deliberations in the penalty phase of a capital case. One of the accepted and embraced roles of the jury is to

\textsuperscript{215} See Sandra Day O’Connor, \textit{Juries: They May Be Broken, But We Can Fix Them}, 44-Jun. Fed. Law. 20, 25 (1997) (“[J]uries are a great institution, with a proud history. As we approach the 21st century, however, we need to make sure we do not remain so wedded to practices hailing from the 20th, or the 18th, or the 13th, that we make it difficult for juries to do their job well. It is my hope that everyone concerned with the proper functioning of our justice system will take this issue seriously, to think hard about ways in which juries can be made to work better, and not to fear change simply because it is different.”).

\textsuperscript{216} See Nunn, \textit{supra} note 203, at 437 n.174 (citing studies showing that at least three minority jurors are required to withstand the pressure of a nine-person racial majority on a jury); Kim Taylor-Thompson, \textit{Empty Votes in Jury Deliberations}, 113 Harv. L. Rev. 1261, 1296 (2000) (suggesting that women jurors may participate at lower rates than men in mixed-gender settings such as juries).

\textsuperscript{217} See, e.g., Lawson v. Borg, 60 F.3d 608, 613 (9th Cir. 1995) (“The number of jurors affected by the misconduct does not weigh heavily in the prejudice calculus for even a single juror’s improperly influenced vote deprives the defendant of an unprejudiced, unanimous verdict.”); United States v. Delaney, 732 F.2d 639, 643 (8th Cir. 1984) (“If a single juror is improperly influenced, the verdict is as unfair as if all were.”) (quoting Stone v. United States, 113 F.2d 70, 77 (6th Cir. 1940)).
act as the conscience of the community and, as such, to exercise mercy. If jurors knew that their deliberations were being recorded for review in case of a death sentence, their tendencies to cross their t’s and dot their i’s would bias the process in favor of death. First, it is often difficult to justify in words why mercy should be exercised when faced with a murderer; mercy is an exercise in human forgiveness. The law does not require such articulation. Second, the practice of checking boxes and articulating reasons tend to favor a decision to impose death. In a capital case, the jury will often find the defendant to be death-eligible after weighing aggravating and mitigating circumstances, but the decision whether to exercise mercy after that point will be discretionary. It is far easier to justify death in words than it is to justify mercy.

Even if this were not the effect of recordings, there is still a legitimate concern that free and open debate will be inhibited. If jurors knew their deliberations were being televised or transcribed, they might not feel as free to state their thoughts or opinions for a variety of reasons: they cannot articulate them well, their ideas are unpopular or make an individual juror sound biased or racist, or they have any of a host of other reasons. Although it is possible that recording deliberations may have the positive effect of making jurors behave more civilly, or may encourage accountability, if the purpose of reviewing

---

218 Similarly, special verdict forms in criminal cases have been criticized as tending to guide a jury toward guilt. See United States v. Spock, 416 F.2d 165, 182 (1st Cir. 1969) (“There is no easier way to reach, and perhaps force, a verdict of guilty than to approach it step by step.”).

219 The effect of televising jury deliberations has been debated. Compare Abramovsky & Edelstein, supra note 132, at 874 (noting that, although jurors in the PBS documentary nullified the conviction and therefore may have seemed unaffected, they exhibited the “Hawthorne effect,” which occurs when “[p]eople aware that they are being observed alter their behavior (in this case, adopting vast eloquence and extreme circumspection) to meet what they imagine to be the expectations of the observers”) (alteration in original), with William R. Bagley, Jr., Note, Jury Room Secrecy: Has the Time Come to Unlock the Door?, 32 Suffolk U. L. Rev. 481, 502 (1999) (“Although the CBS and PBS documentaries provide a limited pool of information from which to draw any absolute conclusions, they demonstrate that jurors will concentrate on the task before them and forget about the presence of a recording devise [sic].”).

220 See Abramovsky & Edelstein, supra note 132, at 892 (“If jury deliberations are routinely recorded, any gain in accountability is offset by the damage to free debate in the jury room . . . .”).

221 See Amar, supra note 214, at 1187 (suggesting videotaping deliberations as educational material); Jackson, supra note 200, at 486 (“If decision makers have to explain their decisions to the community or to some reviewing body, then it is more likely that their decisions can be justified.”); Ruprecht, supra note 132, at 217 (arguing that jury deliberations should be transcribed as part of the record and subject to a limited judicial review).
the deliberations is, in part, to ferret out bias or prejudice, then submerging the expression of those feelings thwarts that goal.

Therefore, although recording jury deliberations has the advantage of providing a true transcript of what was spoken, it will favor the articulate, potentially help spawn a death verdict, and may drive evidence of bias or prejudice underground. The costs of this method outweigh its benefits.

B. Post-Sentence Questioning

Another point in the process where we can determine whether jurors acted arbitrarily in imposing the death penalty is while the jury is still empanelled and just after it is polled in open court as to a unanimous decision for death.\textsuperscript{222} Two distinct methods could be employed: a post-sentence voir dire or a post-sentence questionnaire, either written or oral. However, the drawbacks of these methods lie in the fact that they would occur in the control and presence of institutional pressures to affirm the decision.

The voir dire format would require that each juror be questioned individually,\textsuperscript{223} in the presence of the judge, the prosecution, and the defense. The questions would likely begin with the open-ended question of why the juror voted for death followed by more discrete questions, pinpointing specific areas of concern. The clear benefit of such a format is that it would occur both close in time to the deliberations, when memories are fresh, and in the presence of all parties.

However, the drawbacks stem from those same factors. The timing is problematic because human nature will invariably lead the juror to defend the awesome decision she has just made, rather than expose its weaknesses.\textsuperscript{224} Additionally, the presence of all parties in this

\textsuperscript{222} Although polling is designed to ensure the verdict represents the verdict of each individual juror, it is merely a “yes” or “no” question and is not conducive to discovering arbitrary decision making.

\textsuperscript{223} Individual questioning is crucial for a host of reasons: it permits more honest answers outside the pressure of fellow jurors; the decision to impose death must be the individual’s own decision, and not just the group’s; and jurors are not required to unanimously agree on which mitigating circumstances were proven, see \textit{Mills v. Maryland}, 486 U.S. 367, 384 (1988).

\textsuperscript{224} See \textit{Smith v. Phillips}, 455 U.S. 209, 230 (1982) (Marshall, J., dissenting) (“[G]iven the human propensity for self-justification, it is very difficult ‘to learn from a juror’s own testimony after the verdict whether in fact he was ‘impartial.’ ’ Certainly, a juror is unlikely to admit that he had consciously plotted against the defendant during the course of the trial.” (quoting Phillips v. Smith, 632 F.2d 1019, 1022 (2d Cir. 1980) (citation omitted))), rev’d 455 U.S. 209 (1982); Thompson, \textit{supra} note 136, at 1218 (“The jurors [post-verdict] have a substantial self-interest in providing testimony that is consistent with the validity of
format is not conducive to thoughtful and honest reflection by the juror. In such close quarters with the judge—the symbol of the authority and power of the justice system—the juror will be less likely to reveal anything inopportune. Moreover, the pattern of questioning would very likely imitate that of individual voir dire during jury selection: each party would attempt to coach the juror as to the “right” answer to the question, and the judge would exercise his or her institutional or political pull toward finality.225

The other option at this point in the process is to give each juror a questionnaire after returning a death sentence, to be filled out privately and individually before being excused. The questionnaire would become part of the record on appeal. Compared to the voir dire format, this format has the advantage of occurring outside the intimidating and influential presence of the judge and the lawyers. A disadvantage here is that some jurors may be illiterate, others may be intimidated by the written format, and some jurors will be more articulate than others.

However, the overwhelming problem with this format is, as with post-sentence voir dire, that it suffers from the weakness of timing and location. Again, occurring moments after the jurors have just come to the very difficult decision to send a man to his death, the human urge to defend this decision will be at its zenith. Given the timing and the courthouse setting, the jurors would rightly perceive the questionnaire as asking for an articulation of their reasons, but may not perceive it is as a method to expose any weaknesses in the reasoning.226

The tendency may well be to rubberstamp the proceedings and leave their verdict. They have taken a public position on the issue and may be reluctant to provide testimony that they may have acted improperly in arriving at that position.”)


226 Professor Amar has suggested that if a criminal jury “would like to,” it could explain its reasons for its verdict, perhaps by allowing a clerk assigned to the jury to help “compose a statement of reasons that will enhance public understanding and education.” Amar, supra note 214, at 1187. This suggested reform is not aimed at, nor will it achieve, exposure of the jurors’ errors, caprice, or bias.
the courthouse as quickly as possible. Although that may not always be the case, it must be remembered that the goal here is to find the best method for exposing arbitrariness, not the best method for justifying the verdict.

C. Post-Trial Interviews of Jurors

Post-trial interviews of jurors may be the most reliable method for gathering evidence that exposes arbitrariness in the decision to impose death. The interviews would be conducted outside the intimidating setting of the courthouse and the judge. Although there are distinct disadvantages to this method as well, the criminal justice system is designed to handle these disadvantages.

Under this method, the interviewer will most likely be a defense attorney. Although the Eighth Amendment mandate may benefit the entire criminal justice system as well as the public, the reality is that the only party with a concrete institutional interest in uncovering arbitrariness is the defendant. Therefore, the person most likely to seek out the jurors post-trial is the defense attorney. It is concededly less than ideal to rest the burden of this Eighth Amendment function squarely on the shoulders of the defense attorney. It requires competent and diligent counsel who will take the time, not only to interview the jurors, but to do it well. A competent defense attorney would consult psychologists and death penalty specialists to craft questions that would be most likely to elicit honest and informative responses. Unfortunately, competent defense counsel is often the exception. The reality here is that no reform of the capital system is effective without improvement in the competency levels of capital counsel. This proposal rises or falls dependent upon the availability of competent counsel.

In order to facilitate post-trial interviews and the subsequent introduction of the juror’s testimony into a court of law, courts must remove basic impediments. Because the rules prohibiting contact with jurors and prohibiting juror testimony as to their mental processes lack convincing rationale and conflict with and undermine the Eighth Amendment constitutional mandate, this proposal calls upon courts to find the operation of those rules unconstitutional in capital cases.227

227 Justice Ginsburg has suggested easing the harshness of Rule 606(b) in capital cases:

While precedent supports the Fifth Circuit’s affirmation that statements attesting to the juror’s understanding of the instructions are inadmissible, the statements [the defendant] submitted do assert that apprehension of a lesser
The advantage of post-trial interviews is that they allow the juror the time and space to reflect upon the decision and give an honest assessment of what she thought and did, and of what she observed others saying or doing.\textsuperscript{228} If a juror is going to expose weaknesses in the process, she is more likely to do it here than in the courthouse. However, it is also true that some jurors will take advantage of the opportunity to change their minds or remember things differently.\textsuperscript{229} Defense attorneys are not neutral parties and they will, at the very least, encourage such revelations.

The answer to this problem is the same as it is in any criminal case. The vagaries of human recall never justify forgoing witness testimony. After a post-trial interview, the juror will testify at a hearing and will be subject to cross-examination. Although our adversary system is by no means a model of perfection for finding the truth, it is the closest we have come. Whereas the concern of the prior proposals is that jurors will reveal too little or rubberstamp the decision, here the danger is in overstating or fabricating evidence of arbitrariness. Post-sentence questioning might prevent information from coming before the court, whereas this method more likely involves too much information, both reliable and unreliable, which courts are used to weighing.

What kinds of testimony might the court hear and what can we expect the court to do with it? There are potentially four categories of statements jurors may make. First, there are statements showing the juror to have been mistaken about the law. For example, a juror may have believed that the mitigating factors needed to be proven beyond a reasonable doubt or that, once aggravating factors were found to sentence the judge might impose in fact caused jurors to vote for a death sentence. On a matter so grave, I would not discount those statements altogether.\footnote{Jones v. United States, 527 U.S. 373, 416 n.19 (1999) (Ginsburg, J., dissenting) (citation omitted).}

\textsuperscript{228} Asking jurors what other jurors said or did may be the only way of discovering evidence of racial bias. \textit{See, e.g.}, Taylor-Thompson, \textit{supra} note 216, at 1287–88 (jurors of color can help identify when race is “in play,” using a process described as “going meta,” which “may permit the juror of color to observe the degree to which race affects and infects the jury deliberation process”).

\textsuperscript{229} Short of getting inside a juror’s mind, any interview process is going to have its failings. “[W]hat people say about their own behavior can be very unreliable.” Nunn, \textit{supra} note 203, at 437 n.171. People forget, lie, and exaggerate, as well as tell the truth. See Stephen P. Garvey, \textit{The Emotional Economy of Capital Sentencing}, 75 N.Y.U. L. Rev. 26, 29 (2000) (describing the limitations of juror interviews, such as: answers may be less than forthright, may be what the juror thinks the interviewer wants to hear, or may be the answers believed the most socially acceptable; the juror’s memory may have faded or changed since the time of trial; and the juror may be biased by hindsight).
outweigh mitigating factors, death was mandatory. In these cases, the court must discover what effect the mistake had on the juror’s decision. In order to uphold the death sentence, the state must prove to the court beyond a reasonable doubt that the juror’s mistake did not affect the outcome.\(^\text{230}\) If the juror explains that she would have voted for death in any case, then the mistake did not infect the verdict. But, if the juror says she believes she may have voted for life if she had understood the law correctly, then the death sentence must be vacated.

A second category of statements is a juror’s disclosure that some irrelevant factor, external to the merits of the decision, infected her decision. For example: “I just gave up;” “I wanted to go home;” “The other jurors told me that he would never be executed;” “If I voted for life, he would have been back on the streets in five years;” or “As the only black juror, I felt a lot of pressure to join the others and vote for death.” Again, in this case, the court could affirm the death sentence only if it could find beyond a reasonable doubt that those beliefs did not affect the juror’s vote. In order to determine this, the juror herself would have to testify to its effect, and the prosecution could call other jurors to affirm or deny the existence of any statements made or of external pressures placed upon the juror.

A third category would be statements where a juror described the dynamics or statements of other jurors. Bias and prejudice are most likely to be revealed in this manner.\(^\text{231}\) For example, a juror may say: “Juror X made a number of racist comments about the defendant;” “Some of the jurors voting for death were very threatening toward the two hold-outs for life;” or “Juror Y refused to discuss the mitigating circumstances.” A hearing on the statements would obviously require the testimony of other jurors, and the judge would have to determine the credibility of the testimony. If the testimony leaves reasonable doubt about what happened and its effect on the verdict, then the death sentence must be vacated.

The fourth category of statements would be a juror’s indication that she voted for death based upon a moral response to the situation. For example, a juror may explain that she voted for death because the defendant did not show any remorse, because she was responding to

\(^{230}\) Phillips, 455 U.S. at 219 (stating that claims of juror bias and misconduct are subject to harmless error analysis); Chapman v. California, 386 U.S. 18, 24 (1967) (establishing the harmless error standard of review for errors of constitutional magnitude).

\(^{231}\) Because racism is the factor least likely to be revealed through juror questioning, McCleskey v. Kempp-type statistics should be enough to raise a reasonable doubt that the verdict was based upon improper considerations of race.
the power and pain of the victim impact statement,\textsuperscript{232} or because she simply thought the defendant was a bad man. These responses go to the heart of the debate over our ability to control the moment of the decision to impose a death sentence. We ask the jurors to act as the “conscience of the community” and we desire a “reasoned moral response.”\textsuperscript{233} However, before reaching this point of discretion, a juror must have seriously considered aggravating and mitigating circumstances and reached the decision that the defendant was death-eligible. The guided discretion schemes, while not sufficient, are necessary. After that point, a decision to impose death may be discretionary, but it does require a reasoned moral response.\textsuperscript{234} That response need not be terribly articulate and it need not mimic the aggravating factors, but it cannot be arbitrary. Therefore, as long as the jurors followed the statutory guidelines first, reasons such as those in the statements described above are unlikely to undermine the verdict.

Far from being unimaginable, the proposal here is a familiar process. Courts already hear juror testimony on claims of extraneous influences on the jury’s verdict, and courts already engage in hearings on new evidence during habeas corpus proceedings. With very little cost to the system, claims of arbitrariness can be litigated post-trial. Rather than guessing as to the effect of the arbitrary factor, courts would now be able to turn to the best evidence—the juror’s own statements. Although the reliability of those statements is still a matter for the court to resolve, the court will, for the first time, have competent evidence before it about how the death sentence was determined.

\section*{Conclusion}

For almost thirty years, the Supreme Court has been monitoring capital punishment through a system of risk management. States

\textsuperscript{232} Although the Court has decided otherwise in \textit{Payne v. Tennessee}, 501 U.S. 808, 827 (1991), some have argued that victim impact testimony is an arbitrary factor. \textit{See, e.g., Jeffrey Abramson, Death-is-Different Jurisprudence and the Role of the Capital Jury, 2 Ohio St. J. Crim. L.} 117, 134 (2004) (“Any message that weighing of the relative value of lives is germane to the penalty decision is especially alarming in light of statistical studies showing that a victim’s high socio-economic status seems to touch off an ‘invisible bias’ in sentencing authorities . . . .”).

\textsuperscript{233} Indeed, this is why, in capital cases, juries, and not judges, may be the preferred sentencers. \textit{See Ring v. Arizona}, 536 U.S. 584, 614–15 (2002) (Breyer, J., concurring) (arguing that jurors are better situated to express or represent the moral sensibilities of the community).

adopted guided discretion statutes on the direction from the Court that, as long as they put in place procedures designed to reduce a substantial risk of arbitrary imposition of death, then the Eighth Amendment was satisfied. According to the Court, it is of no constitutional moment whether these procedures actually work. This Article has demonstrated that Furman v. Georgia’s mandate is absolute: the Eighth Amendment commands a nonarbitrary outcome when inflicting the gravest punishment known to civilized society.235

We need to move from procedures that shield arbitrariness to those that expose it. The only procedure that comes close to enforcing the substantive right to a nonarbitrary verdict is review of the jurors’ reasons for imposing death. It is no longer sufficient to raise the flag of jury secrecy against such procedures. Post-trial practices enforcing jury secrecy are difficult to justify in capital cases. Although there are practical hurdles to realizing the goal of jury transparency, they are not insurmountable. It requires no imagination or effort from courts to abolish the no-impeachment rule in capital cases and hold post-trial hearings.

Perhaps abolition of the capital punishment system is the only tolerable solution to its fallibilities.236 However, abolition is not imminent. In the meantime, human beings face death in the darkness of ignorance. They, and we, deserve to know, and the Constitution demands knowing, whether they await death because of any prejudice, error, or caprice operating among the jurors charged with the awesome responsibility of deciding if they will live or die. Viewing, and reviewing, the capital jurors’ reasons for imposing death shines light into that dark space.237


236 For example, both racism and innocence present uniquely intractable problems that call for an end of capital punishment. See Howe, supra note 102, at 2145–49, 2165–66 (arguing that racial discrimination in the death penalty is impossible to eliminate and violates the Eighth Amendment); Radin, supra note 74, at 1184 (“The issue is really whether we can accord due respect to any defendant sentenced to death in the context of a system that we know must wrongly kill some of them although we do not know which.” (citation omitted)); Richard A. Rosen, Innocence and Death, 82 N.C. L. Rev. 61, 106 (2003) (arguing that preventing the execution of innocents is impossible and therefore capital punishment is unconstitutional).

237 See Carol S. Steiker & Jordan M. Steiker, Should Abolitionists Support Legislative “Reform” of the Death Penalty?, 63 OHIO ST. L.J. 417, 427–28 (2002) (questioning the propriety of reform while advocating as a “non-entrenching” reform, “sunshine reform, or more simply, data collection and dissemination,” such as the collection of prosecutors’ reasons for bringing or declining to bring a capital case).
STRUCTURE VERSUS EFFECT: REVEALING THE UNCONSTITUTIONAL OPERATION OF TITLE IX’S ATHLETICS PROVISIONS

Abstract: Title IX of the 1972 Education Amendments prohibits discrimination on the basis of sex in any educational program receiving federal financial assistance, including athletics. Although the statute has brought great improvements in female athletics, such achievements have not come without problems. Title IX’s complex regulatory framework, developed by the Department of Education, has led to numerous lawsuits. In early cases, courts interpreted and upheld the framework as consistent with Title IX. This Note argues, however, that the courts’ early decisions and interpretation of the framework, coupled with educational institutions’ current fiscal constraints, have left institutions with only one option for compliance: cutting men’s teams. Such an inflexible framework amounts to a gender-based quota system that, under recent affirmative action caselaw, courts should hold unconstitutional. Only then will the Department of Education have the impetus to reinterpret the regulatory framework so that Title IX can operate in the way in which it was originally intended.

Introduction

Congress enacted Title IX as part of the Education Amendments of 1972 in response to its finding of pervasive discrimination against women with respect to educational opportunities. Title IX’s objectives are to prevent the use of federal funds to support discriminatory practices and to provide individual citizens protection against those practices. Despite its seemingly clear language, however, application of Title IX has not come without problems. Indeed, the complex regulatory framework that the Department of Education (the “DOE”)
developed in an effort to clarify the statute’s athletics provisions has resulted in numerous lawsuits and the subsequent need for judicial interpretation.\(^4\)

In 1993, in *Cohen v. Brown University*, the United States Court of Appeals for the First Circuit heard the first challenge brought by a female under the Title IX athletics provisions.\(^5\) This case presented the first opportunity for the federal court system to analyze and interpret the complex regulatory framework that governs the enforcement of Title IX.\(^6\) The court’s conclusions were pivotal because virtually every federal circuit court of appeals that has subsequently ruled on a Title IX athletics discrimination case has followed the First Circuit’s analysis.\(^7\)

This deference to the *Cohen* court’s analysis was justified initially because most challenges under Title IX’s athletics provisions were characteristically similar to *Cohen*, in which female athletes were challenging a school action that had an adverse effect on their opportunity to participate.\(^8\) More recently, however, the characteristic challenge under Title IX’s athletics provisions has changed, with the challenge instead coming from male athletes.\(^9\) These athletes contend that, under the current interpretation of Title IX’s regulatory framework, universities faced with mandatory budget cuts are left with only one option for complying with Title IX: cutting men’s teams.\(^10\) Because such decisions are based solely on gender, they violate Title IX and the Fourteenth Amendment.\(^11\) Thus far, however, the courts have approved these cuts as being consistent with the mandate of Ti-

---

\(^4\) See, e.g., *Miami Univ. Wrestling Club*, 302 F.3d at 613–16; *Pederson*, 213 F.3d at 864, 876–82; *Cohen I*, 991 F.2d at 892–93.

\(^5\) *Cohen I*, 991 F.2d at 891.

\(^6\) Id.


\(^8\) See, e.g., *Pederson*, 213 F.3d at 864; *Horner*, 43 F.3d at 268; *Roberts*, 998 F.2d at 826.


\(^10\) See *Miami Univ. Wrestling Club*, 302 F.3d at 610; *Boulahanis*, 198 F.3d at 639; *Kelley*, 35 F.3d at 267.

\(^11\) See *Miami Univ. Wrestling Club*, 302 F.3d at 615; *Boulahanis*, 198 F.3d at 639; *Kelley*, 35 F.3d at 267.
tle IX, and at least one court has determined that these cuts do not offend the Fourteenth Amendment. In reaching this conclusion, the courts have continued to rely on the analysis of the First Circuit in Cohen.

This Note argues, however, that such continuing deference to the First Circuit’s analysis and reasoning in Cohen, in light of the substance of the recent male challenges to Title IX and the fiscal realities in higher education, is improper, and it is time for the circuit courts to alter their analysis of these claims. The courts today are faced with a set of claims—those based on the consequences of the Cohen decision in light of the reality of fiscal constraints—that are different from those faced by the Cohen court, and, as such, it is proper for the courts to reevaluate the original interpretation the First Circuit gave to Title IX’s regulatory framework. Examination of how this original interpretation has operated in practice since the 1993 decision reveals an inherently inflexible structure that has resulted in decision making based solely on gender. This inflexibility, particularly when considered in light of the U.S. Supreme Court’s recent 2003 affirmative action rulings in Grutter v. Bollinger and Gratz v. Bollinger, should encourage the circuit courts to determine that the regulatory framework of Title IX’s athletics provisions is unconstitutional.

Part I of this Note provides an overview of the legislative history of Title IX and the development of the regulatory framework that governs the enforcement of its athletics provisions. This Part then discusses the first judicial interpretation of Title IX’s athletics provisions in Cohen, including the First Circuit’s important analysis and conclusions regarding the complex regulatory framework. Although virtually every circuit court has followed this analysis, this Part also

---

12 Chalenor, 291 F.3d at 1043, 1048–49 (approving the elimination of men’s wrestling); Kelley, 35 F.3d at 269, 272–73 (approving the elimination of diving, fencing, and men’s swimming while retaining women’s swimming); see Miami Univ. Wrestling Club, 302 F.3d at 611, 615–16 (approving the elimination of men’s soccer, tennis, and wrestling).

13 E.g., Neal, 198 F.3d at 771; Boulahanis, 198 F.3d at 639; Horner, 43 F.3d at 275; Roberts, 998 F.2d at 830 & n.9.

14 See infra notes 254–74 and accompanying text.

15 See Miami Univ. Wrestling Club, 302 F.3d at 610, 615; Boulahanis, 198 F.3d at 639; Kelley, 35 F.3d at 267, 272.

16 See Miami Univ. Wrestling Club, 302 F.3d at 611; Chalenor, 291 F.3d at 1043; Kelley, 35 F.3d at 269.


18 See infra notes 29–55 and accompanying text.

19 See infra notes 56–91 and accompanying text.
discusses two district court rulings that offer concerns about the First Circuit’s approach.20 Part II provides a discussion of the more recent male challenges to Title IX’s athletics provisions.21 It highlights the realities of fiscal pressures on educational institutions and how the resulting necessity for budget cuts, coupled with the Cohen court’s interpretation of Title IX’s regulatory framework, has forced institutions to cut only men’s teams.22 In light of the nature of the male athletes’ claims, Part III suggests a correlation between the male Title IX claims and the claims in affirmative action cases, and thus provides an overview of the Supreme Court’s current thought on the issue of affirmative action programs in higher education.23

Finally, Part IV critically analyzes the First Circuit’s original interpretation of Title IX’s regulatory framework and rejects it as creating a framework that is impermissibly inflexible under the Equal Protection Clause of the Fourteenth Amendment.24 Part IV.A explains the various flaws in the First Circuit’s reasoning in Cohen and concludes that the court’s deference to the DOE’s original Policy Interpretation of Title IX was inappropriate, and that its interpretation of prong three of the DOE’s three-part test and of the operation of the test as a whole was flawed.25 Part IV.B then reveals that the First Circuit’s interpretation has in effect resulted in the elimination of any flexibility the regulatory framework originally possessed.26 Part IV.C argues that this lack of flexibility is directly at odds with the Supreme Court’s current affirmative action jurisprudence, which emphasizes the need for flexibility when decisions are made based on an individual’s membership in a particular class.27 Part IV.D, finally, summarizes the foregoing analysis and suggests that the circuit courts alter their analyses of Title IX claims and instead hold that the regulatory framework, as currently interpreted, is unconstitutional.28

---

20 See infra notes 92–104 and accompanying text.
21 See infra notes 111–52 and accompanying text.
22 See infra notes 111–20 and accompanying text.
23 See infra notes 153–80 and accompanying text.
24 See infra notes 181–274 and accompanying text.
25 See infra notes 186–225 and accompanying text.
26 See infra notes 226–39 and accompanying text.
27 See infra notes 240–53 and accompanying text.
28 See infra notes 254–74 and accompanying text.
I. TITLE IX: LEGISLATIVE HISTORY AND JUDICIAL INTERPRETATION

A. Legislative History of Title IX

Before examining the problems with Title IX’s regulatory framework, it is first necessary to understand how this complex framework came into being.  Although the objective of the statute—to prevent the use of federal funds to support discriminatory practices in education—was clear, the language of the statute and the subsequent ambiguity in its application to athletics resulted in the development of a complicated regulatory framework through which the statute is now applied.  The courts rely on this complex regulatory scheme to guide their interpretation of Title IX’s athletics provisions.

1. Statutory and Regulatory Framework

The language of the statute makes clear that Title IX prohibits gender discrimination in educational programs or activities receiving federal financial assistance.  It provides, in relevant part in § 1681, that no person shall be discriminated against on the basis of sex in any federally funded education program or activity.  Also in § 1681, the statute specifically provides that statistical evidence of disparities between genders in education programs and activities cannot alone lead to a finding of discrimination.


33 Id. The drafters modeled Title IX’s general prohibition against sex discrimination after Title VI of the Civil Rights Act of 1964.  N. Haven Bd. of Educ. v. Bell, 456 U.S. 512, 514 (1982).  Title VI provides that “[n]o person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.”  42 U.S.C. § 2000d (2000).

34 20 U.S.C. § 1681(b).  In particular, the statute states that

[n]othing contained in . . . this section shall be interpreted to require any educational institution to grant preferential or disparate treatment to the members of one sex on account of an imbalance which may exist with respect to the total number or percentage of persons of that sex participating in or receiving the benefits of any federally supported program or activity, in com-
In connection with the passage of Title IX, Congress did not produce any secondary legislative materials; there was no committee report included with the bill. This lack of legislative history, along with no specific mention of athletics in the statute itself, rendered ambiguous the application of Title IX to intercollegiate athletics. In response to growing confusion over Title IX’s application to intercollegiate athletics, Congress directed the DOE, through its Office of Civil Rights (“OCR”), to develop a regulation implementing the provisions of Title IX, including specific provisions for intercollegiate athletics. Pursuant to this directive, the DOE promulgated a Final Regulation (the “Regulation”) in 1975, which remains in effect.

The Regulation mirrors Title IX’s language by generally prohibiting discrimination of student-athletes based on sex. In addition to this general provision, the Regulation addresses federally funded athletics programs and provides that a recipient of federal funds shall provide equal athletic opportunities for members of both sexes. The Regulation then lists ten factors to be used to assess equivalence between male and female athletic opportunities. Factor one involves a consideration as to whether the selection of sports and levels of comparison with the total number or percentage of persons of that sex in any community, State, section, or other area.

Id.

35 Cohen I, 991 F.2d at 893; Connolly & Adelman, supra note 30, at 850 & n.17.
36 Connolly & Adelman, supra note 30, at 850. Congressional records indicate that intercollegiate athletics were only mentioned twice during the congressional debate on the statute. 118 Cong. Rec. 5807 (1972) (statement of Sen. Bayh on privacy in athletic facilities); 117 Cong. Rec. 30,407 (1971) (statement of Sen. Bayh noting that the proposed Title IX would not require mixed-gender football teams).
37 Education Amendments of 1974, Pub. L. No. 93–380, 88 Stat. 484, 844 (1974). Agency responsibility for the administration of Title IX shifted from the Department of Health, Education, and Welfare (“HEW”) to the DOE when HEW was split into two agencies, the DOE and the Department of Health and Human Services. See Pub. L. No. 96–88, 93 Stat. 668 (1979). The regulations and the agency documents discussed in this Note were originally promulgated by HEW, the administering agency at the time, and later adopted by the present administrative agency, the DOE. See Cohen I, 991 F.2d at 895. For the sake of simplicity, this Note will treat the DOE as the promulgating agency.
38 34 C.F.R. § 106.1–.71 (2004). The promulgation of the Final Regulation followed a four-month period in which the DOE received and considered over 9700 comments, suggestions, and objections to the proposed regulation. Nondiscrimination on the Basis of Sex in Education Programs and Activities Receiving or Benefiting from Federal Financial Assistance, 40 Fed. Reg. 24,127, 24,128 (June 4, 1975) (codified at 34 C.F.R. § 106.1–.71 (2004)) [hereinafter “Nondiscrimination on the Basis of Sex”].
39 20 U.S.C. § 1681(a); 34 C.F.R. § 106.41(a).
40 34 C.F.R. § 106.41(c).
41 Id.
petition effectively accommodate the interests and abilities of both sexes; factors two through ten consider equivalence in the provision of other services, facilities, and resources.42

2. The DOE’s 1979 Policy Interpretation

After the DOE issued the Regulation, however, confusion followed as to how properly to apply the ten factors and otherwise comply with Title IX.43 As a result, the agency in 1979 issued a Policy Interpretation, which contains three separate sections.44 Section one clarifies the statute’s application with respect to scholarships; section two discusses the use of factors two through ten of the ten-factor test; and the third section sets forth an interpretation of factor one—whether the institution effectively accommodates the interests and abilities of both sexes.45

Under this third section, the DOE provides that institutions may use “any non-discriminatory” method to assess students’ athletic interests and abilities, and further offers guidance as to appropriate mechanisms for assessing such interest.46 This section then sets forth a three-part test to determine if the institution is effectively accommodating the interests and abilities of both male and female athletes: (1) whether

42 Id. Specifically, factors two through ten were as follows:

(2) [t]he provision of equipment and supplies; (3) [s]cheduling of games and practice time; (4) [t]ravel and per diem allowance; (5) [o]portunity to receive coaching and academic tutoring; (6) [a]ssignment and compensation of coaches and tutors; (7) [p]rovision of locker rooms, practice and competitive facilities; (8) [p]rovision of medical and training facilities and services; (9) [p]rovision of housing and dining facilities and services; (10) [p]ublicity.

Id. The OCR stated that this list was intended to serve as guidance on what factors the DOE considered integral to providing equal athletic opportunities. Nondiscrimination on the Basis of Sex, supra note 38, at 24,134. The enumerated items were not intended to be an exhaustive list. See id.

43 See Title IX of the Education Amendments of 1972; A Policy Interpretation, 44 Fed. Reg. 71,413, 71,413 (Dec. 11, 1979) [hereinafter “Policy Interpretation”]. By November 1978, the DOE had received almost one hundred complaints alleging that more than sixty-two colleges and universities were violating Title IX by not providing women equal opportunities in athletics. 43 Fed. Reg. 58,070, 58,071 (Dec. 11, 1978).


45 Id. at 71,414.

46 Id. at 71,417. The Policy Interpretation requires that this nondiscriminatory method take into account the nationally increasing levels of women’s interests and abilities, that the method of determining interest not disadvantage the members of the underrepresented sex, that the method of determining ability take into account team performance records, and that the methods used generally be responsive to the expressed interests of students of the underrepresented sex who are capable of participating in intercollegiate competition. Id.
participation opportunities are provided in numbers “substantially proportionate to their respective enrollments,” (2) “whether the institution can show a history and continuing practice of program expansion,” and (3) whether the interests and abilities of the members of the underrepresented sex have been fully and effectively accommodated. Proper interpretation and application of this three-part test has been the subject of most, if not all, of the Title IX athletics litigation.

3. The DOE’s 1996 Policy Clarification

Guided by the Policy Interpretation, universities continued to attempt to comply with Title IX, but these attempts at compliance created a new set of concerns among institutions, athletes, and members of Congress. In June 1995, after holding a hearing on Title IX and the three-part test outlined above, 142 members of Congress wrote to the DOE expressing concern over the fact that educational institutions were satisfying prong one of the three-part test—substantial proportionality of opportunity—by eliminating men’s athletic opportunities. In response to this request for more guidance about the existing standards governing the enforcement of Title IX, the DOE in 1996 issued a Policy Clarification.

The Policy Clarification provides specific factors intended to guide an analysis under each part of the three-part test and includes examples to demonstrate how these factors should be analyzed in practice. It underscores the fact that the requirement addressed by the three-part test—the effective accommodation of interests and abilities of both sexes—is only one of many factors considered under the Regulation to determine if an institution is in compliance with

47 Id. at 71,418.
48 See, e.g., Kelley, 35 F.3d at 267–68, 270–72; Cohen I, 991 F.2d at 893–900.
49 See Nat’l Wrestling Coaches Ass’n, 263 F. Supp. 2d at 92.
50 Id.
51 Office for Civil Rights, U.S. Dep’t of Educ., Clarification of Intercollegiate Athletics Policy Guidance: The Three-Part Test (Jan. 16, 1996), http://www.ed.gov/about/offices/list/ocr/docs/clarific.html [hereinafter “Policy Clarification”]. Prior to issuing this Policy Clarification, the DOE solicited public comment with respect to the narrow question of whether the Policy Clarification provided the appropriate clarity to areas of the three-part test that had generated questions. Dear Colleague Letter from Norma V. Cantú, Assistant Sec’y, Office for Civil Rights, U.S. Dep’t of Educ. (Jan. 16, 1996), http://www.ed.gov/about/offices/list/ocr/docs/clarific.html [hereinafter “Dear Colleague Letter”]. The DOE received and considered over 200 public comments prior to the issuance of the final Policy Clarification in January 1996. Id.
52 Policy Clarification, supra note 51.
Title IX. Further, the letter accompanying the Policy Clarification describes the first prong as a “safe harbor” for establishing compliance with the equal opportunity requirement. Finally, the Policy Clarification emphasizes that the three-part test provides an institution with three different avenues of compliance, that institutions have flexibility in deciding how to provide nondiscriminatory participation opportunities for their students, and that the OCR does not require quotas.

B. Judicial Interpretation of the Title IX Athletics Provisions

Courts must rely on this complex regulatory framework for guidance when deciding athletes’ Title IX challenges against universities. The common challenge to the athletics provisions of Title IX involves women, as the underrepresented sex, challenging some school action that has an adverse effect on their opportunity to participate. Typically, the challenged action involves either the downgrading of women’s teams from varsity to club status (with its corresponding loss of funding, respect, and ability to attract talented athletes and coaches), the dropping of women’s teams altogether, or the refusal to upgrade women’s club teams to varsity status. In the face of such a claim of discrimination under Title IX, institutions typically defend themselves by challenging the validity of the Regulation and the Policy Interpretation—especially the three-part test. In particular, schools often argue that the Policy Interpretation and its three-part test are inconsistent with Title IX, exceed agency authority, and violate the Equal Protection Clause because they operate to discriminate against men, the overrepresented sex. Much to the dismay of the institutions, however, Title

---

53 Id. The letter was likely alluding to factors two through ten of the Regulation. See id.
54 Dear Colleague Letter, supra note 51.
55 Policy Clarification, supra note 51. In arguing that the three-part test does not operate as a quota, the letter emphasizes the flexibility that the test affords, such that if an institution decides to comply with the effective accommodations test by satisfying prong three, it would not be required to achieve substantial proportionality or show a history of continued expansion of opportunities for the underrepresented sex. Id.
56 See, e.g., Miami Univ. Wrestling Club v. Miami Univ., 302 F.3d 608, 610–13 (6th Cir. 2002); Kelley, 35 F.3d at 270; Cohen I, 991 F.2d at 893–94.
57 See e.g., Pederson v. La. State Univ., 213 F.3d 858, 864 (5th Cir. 2000); Horner v. Ky. High Sch. Athletic Ass’n, 43 F.3d 265, 268 (6th Cir. 1994); Roberts v. Colo. State Bd. of Agric., 998 F.2d 824, 826 (10th Cir. 1993); Cohen I, 991 F.2d at 892–93.
58 See Roberts, 998 F.2d at 826; Cook v. Colgate Univ., 992 F.2d 17, 18 (2d Cir. 1993); Cohen I, 991 F.2d at 892; Connolly & Adelman, supra note 30, at 868.
59 Nat’l Wrestling Coaches Ass’n, 263 F. Supp. 2d at 95.
60 Id.
IX, its Regulation, and the Policy Interpretation thus far have survived constitutional challenges in the federal circuit courts of appeals.\textsuperscript{61} The courts have chosen to give substantial deference to the Regulation and the Policy Interpretation as valid mechanisms for the enforcement of Title IX.\textsuperscript{62}

This consistent and substantial deference to the Regulation and the Policy Interpretation results from the almost universal reliance on the U.S. Court of Appeals for the First Circuit’s analysis in the 1993 watershed case, \textit{Cohen v. Brown University}.\textsuperscript{63} Indeed, virtually every circuit court that has subsequently ruled on a Title IX discrimination in athletics case has followed the First Circuit’s analysis of and level of deference to the Regulation and the Policy Interpretation.\textsuperscript{64}


In 1993, the U.S. Court of Appeals for the First Circuit became the first federal court of appeals to rule on a case involving the athletics provisions of Title IX in the case \textit{Cohen v. Brown University} ("\textit{Cohen I}").\textsuperscript{65} In that case, the plaintiffs, representing a class of current and future women athletes, challenged Brown University’s decision to drop women’s volleyball and gymnastics from intercollegiate varsity status, which Brown had undertaken in an effort to reduce its financial burden.\textsuperscript{66} The plaintiffs alleged that Brown violated Title IX’s ban on gender-based discrimination in its provision of athletics.\textsuperscript{67} They further ar-

\textsuperscript{61} See Miami Univ. Wrestling Club, 302 F.3d at 613–16; Chalenor v. Univ. of N.D., 291 F.3d 1042, 1047–50 (8th Cir. 2002); Pederson, 215 F.3d at 878–82; Neal v. Bd. of Trs., 198 F.3d 763, 772–73 (9th Cir. 1999); Boulahanis v. Bd. of Regents, 198 F.3d 633, 637–39 (7th Cir. 1999); Kelley, 35 F.3d at 271–72; Cook, 992 F.2d at 18; Cohen I, 991 F.2d at 895–900, 905–06.

\textsuperscript{62} See e.g., Neal, 198 F.3d at 771–73; Kelley, 35 F.3d at 271; Roberts, 998 F.2d at 828; Cohen I, 991 F.2d at 899.


\textsuperscript{64} See Miami Univ. Wrestling Club, 302 F.3d at 615; Chalenor, 291 F.3d at 1047; Neal, 198 F.3d at 767; Boulahanis, 198 F.3d at 638; Cohen II, 101 F.3d 155, 172–73 (1st Cir. 1996); Horner, 43 F.3d at 271, 272–73, 275; Roberts, 998 F.2d at 830 & n.9, 831–32; Williams v. Sch. Dist. of Bethlehem, 998 F.2d 168, 171 (3d Cir. 1993); Lamber, \textit{supra} note 63, at 169.

\textsuperscript{65} See Cohen I, 991 F.2d at 891, 907. The First Circuit numbers its \textit{Cohen} cases in a different manner from this Note, but because this Note is only discussing two opinions, it refers to the First Circuit’s initial opinion as \textit{Cohen I} and the First Circuit’s opinion after remand to the district court as \textit{Cohen II}.

\textsuperscript{66} Id. at 892. As part of this plan, Brown also dropped men’s golf and men’s water polo from varsity status. \textit{Id}.

\textsuperscript{67} Id. at 893. Although the elimination of these four teams did not significantly impact the overall participation rate of either gender, the existing participation rates did not rep-
gued that Brown’s decision to devalue two women’s programs without first taking steps to equalize the athletic opportunities, either by reducing men’s activities or by adding other women’s teams to compensate for the loss, further exacerbated this violation.68

After engaging in an overview of the legislative and regulatory history of Title IX, the First Circuit held that the DOE’s Title IX Regulation should be accorded appreciable deference.69 In reaching this conclusion, the court relied on *Chevron U.S.A. Inc. v. Natural Reserve Defense Council, Inc.*, the U.S. Supreme Court’s 1984 decision addressing the issue of judicial deference to administrative agency determinations.70 The court granted the Regulation particularly high deference because Congress had explicitly delegated to the agency the responsibility of setting forth Title IX standards for athletic programs.71 The court then went on to hold that the Policy Interpretation should also be given substantial deference, as it was a reasonable interpretation of the Regulation.72

Having granted deference to the Policy Interpretation, the court then reviewed each of the three sections of the Policy Interpretation—athletic financial assistance, equivalence of other program areas, and effective accommodation of student interests and abilities.73 The court concluded that an institution can violate Title IX even if it complies with the first two sections.74 In other words, under this interpretation, an institution must comply with the third section and its three-part test to prevail in a Title IX challenge.75

The court reviewed the three-part test and accepted the statement in the Policy Clarification that the first benchmark provides a safe harbor for institutions to achieve compliance.76 The court then

---

68 Id. at 893.
69 Id. at 893–95.
71 *Id.* (citing *Chevron*, 467 U.S. at 844 (holding that where Congress has explicitly delegated responsibility to an agency, the agency’s regulations deserve “controlling weight unless they are arbitrary, capricious, or manifestly contrary to the statute” at issue)).
72 Id. at 896–97.
73 Id. at 897.
74 Id.
75 See *Cohen I*, 991 F.2d at 897.
76 Id.
went on to describe that the second and third parts of the test provide additional means for achieving compliance; indeed, the court recognized that it was unlikely that athletic establishments at most institutions would reflect the gender balance of their student bodies.\textsuperscript{77} The court cautioned, however, that the third part of the test sets a high standard in that it requires full and effective accommodation, and that if there is sufficient interest and ability among members of the underrepresented sex that existing programs do not satisfy, an institution necessarily fails the third prong of the test.\textsuperscript{78}

In reaching its holding on the proper application of the third part of the effective accommodation test, the court rejected Brown’s argument that, to the extent that students’ interests in athletics are disproportionate by gender, the third part of the test should be read to allow institutions to meet the interests of students incompletely, so long as the school’s response is in direct proportion to the comparative levels of interest.\textsuperscript{79} In the court’s opinion, Brown’s interpretation “read[] the ‘full’ out of the duty to accommodate ‘fully and effectively.’”\textsuperscript{80} Additionally, the court held that Brown’s view, which would have required the assessment of relative interests and abilities of the student population, would have made it overly difficult for institutions to assess compliance and would have presented significant and difficult quantification problems.\textsuperscript{81} Rather, the court preferred a simpler and easier-to-administer interpretation of prong three, which requires an academic institution to establish a new team or upgrade an existing club team if there is a sufficiently high unmet need in the underrepresented gender.\textsuperscript{82} Thus, the court adopted this high standard primarily based on its ease of administration and not necessarily because of its consistency with agency intent.\textsuperscript{83}

\begin{itemize}
\item \textsuperscript{77} Id. at 897–98.
\item \textsuperscript{78} Id. at 898.
\item \textsuperscript{79} Id. at 899–900. The court provided an example of Brown’s reasoning: suppose a university, “Oooh U.”, has a student body consisting of 1000 men, 500 of whom are interested and able to compete, and 1000 women, 250 of whom are interested and able to compete. \textit{Id.} at 899. Under Brown’s reasoning, Oooh U. would have had to provide athletic opportunities in line with the two-to-one interested athletes ratio—here, 100 slots for men and fifty slots for women. \textit{Id.}
\item \textsuperscript{80} \textit{Cohen I}, 991 F.2d at 899 (quoting Policy Interpretation, supra note 43, at 71,418). Continuing its hypothetical, the court held that to satisfy prong three, Oooh U. would have had to accommodate the interests of all 250 women. \textit{Id.}
\item \textsuperscript{81} Id. at 900.
\item \textsuperscript{82} Id.
\item \textsuperscript{83} See id.
\end{itemize}
Having reached the above conclusions, the First Circuit remanded the case to the district court for further proceedings on whether Brown had violated Title IX. On remand, the district court applied the test laid out in Cohen I and determined that Brown’s athletic programs violated Title IX. Brown then appealed the decision, challenging on constitutional and statutory grounds the test employed by the district court.

In upholding the district court’s findings after remand, the First Circuit, in Cohen II, upheld the Cohen I analysis and conclusions. The court also addressed the claim that its Cohen I test for compliance with Title IX effectively renders Title IX an “affirmative action statute” that results in preferential treatment for women because it imposes quotas that exceed women’s relative interests and abilities in athletics. In rejecting this claim, the court maintained that Title IX is not an affirmative action statute, but rather an anti-discrimination statute, and that no aspect of the Title IX regime mandates gender-based preferences or quotas. Indeed, the court pointed out that the substantial proportionality prong of the three-part test is merely a starting point for the analysis, and that it is but one aspect of the inquiry into whether an institution’s athletics program complies with Title IX. In coming to this conclusion, however, the court failed to consider the fact that its interpretation of prong three had effectively left prong one as the only mechanism for a financially strapped institution to achieve compliance.

2. District Court Concerns with Cohen Deference and Interpretation

Although most courts have followed the analysis and interpretation of the First Circuit in Cohen I and Cohen II, not all have done so without reservation. Indeed, several federal district courts have

84 Id. at 907.
85 Cohen II, 101 F.3d at 161.
86 Id. at 162.
87 Id. at 172–73.
88 Id. at 169.
89 Id. at 170.
90 Cohen II, 101 F.3d at 171.
91 See id. at 170–71. But see Miami Univ. Wrestling Club, 302 F.3d at 613 (recognizing that “in the real world of finite resources, [cutting men’s teams] may be the only way for an education institution to comply with Title IX”).
called into question the substantial deference granted to the Policy Interpretation, as well as the use of proportionality (prong one of the three-part test) as a measure of compliance. These concerns represent valid issues with the current interpretation of Title IX and its regulatory framework.

The U.S. District Court for the Middle District of Louisiana articulated the most explicit rejection of the First Circuit’s analysis and interpretation in its 1996 decision, *Pederson v. Louisiana State University.* In that case, the court pointed out that no President has ever approved the Policy Interpretation, and thus it does not have the binding effect that would normally be accorded to such rules and regulations. Further, and more importantly, the court held that prong one of the three-part test—substantial proportionality—should not be considered a “safe harbor” because stopping the inquiry at the point of numerical equality would not comport with Title IX’s mandate. The court held that although the Policy Interpretation serves as a useful guide, to the extent that it suggests that numerical proportionality between the sexes will suffice, it is contrary to the language of Title IX and should not be followed.

Two additional district courts have expressed reservations about the Policy Interpretation’s proportionality test. In the 1994 case *Kelley v. Board of Trustees,* in which the plaintiffs—male swimmers—sought a preliminary injunction to prevent the university from cutting their program while leaving the female swimming program intact, the U. S. District Court for the Central District of Illinois expressed some con-

---

93 See *Neal District,* 1999 WL 1569047, at *3; *Pederson District,* 912 F. Supp. at 910, 913–14; *Kelley District,* 832 F. Supp. at 241–43.
94 See *Neal District,* 1999 WL 1569047, at *3; *Pederson District,* 912 F. Supp. at 910, 913–14; *Kelley District,* 832 F. Supp. at 241–43.
95 See *912 F. Supp. at 913–14.*
96 Id. at 910.
97 Id. at 913–14. The court pointed out that the assumption underlying the substantial proportionality test is that the interest and ability to participate in sports is equal between all men and women on all campuses. *Id.* at 913. The court stated, however, that no basis exists for such a pivotal assumption. *Id.* Rather, the court believed it more likely that interests and abilities will vary from campus to campus. *Id.* at 914. The court further pointed out that the *Cohen court* (and others) reached the conclusion that the first prong should be a “safe harbor” solely as a result of administrative deference. *Id.*
98 Id. at 914. The court went on to hold that because the university had failed to satisfy prongs two and three of the three-part test, it was in violation of Title IX, regardless of the court’s holding on the validity of prong one. *Id.* at 915–17.
99 See *Neal District,* 1999 WL 1569047, at *3; *Kelley District,* 832 F. Supp. at 241.
cern that the judicial and policy interpretations had essentially “converted Title IX from a statute which prohibits discrimination on the basis of sex . . . into a statute which provides ‘equal opportunity for members of both sexes.’”100 The court also explicitly recognized that the law could have been interpreted differently and that Congress probably did not anticipate that Title IX would produce such “draconian” results whereby men’s teams are eliminated to achieve Title IX compliance.101

With a similar line of reasoning, in 1999 in Neal v. Board of Trustees, the U.S. District Court for the Eastern District of California granted a preliminary injunction to male wrestlers who asserted Title IX claims.102 In issuing the preliminary injunction in favor of the male athletes, the court concluded that relying on proportionality to achieve compliance constituted an implementation of a quota based on gender, and that such a gender-based distinction was a violation of Title IX.103 Although the Ninth, Fifth, and Seventh Circuit Courts of Appeals rejected each of these expressed district court concerns, such concerns reveal the lack of unanimous support for the Cohen court’s interpretation of the regulatory framework.104

3. Christensen v. Harris County: A Revised View on Deference Due to Agency Interpretations

In addition to the district courts’ concerns with the Cohen court’s interpretation, a recent U.S. Supreme Court case also calls into question the deference that the Cohen court granted to the Policy Interpretation.105 In holding that the Policy Interpretation should be granted substantial deference, the Cohen court relied primarily upon Chevron.106 In 2000, however, the Supreme Court, in Christensen v. Harris County, clarified Chevron’s applicability to agency interpretations other than regulations.107 In particular, the Court held that agency interpre-

---

100 832 F. Supp. at 241.
101 Id. at 242, 243. Despite these concerns, the court held in favor of the university and upheld the cuts to the men’s swimming program, reasoning they were consistent with current judicial and agency interpretation of Title IX. Id. at 242, 244.
102 1999 WL 1569047, at *1, *3.
103 Id. at *3.
106 Cohen I, 991 F.2d at 895 (citing Chevron, 467 U.S. at 844).
107 See Christensen, 529 U.S. at 587 (citing Chevron, 467 U.S. at 844).
tations such as those in policy statements, opinion letters, agency manuals, and enforcement guidelines—all of which do not have the force of law—are not entitled to *Chevron*-style deference. The Court stated that such interpretations are entitled to respect only to the extent that they are persuasive. Because the *Cohen* court acted four years before *Christensen*, the First Circuit lacked *Christensen*’s guidance when it accorded substantial deference to the Policy Interpretation.

II. Recent Challenges to Title IX: Discrimination Against Men

In contrast to the typical early challenges to Title IX brought by female athletes, currently the more frequent challenges to the statute come from male athletes, who contend that the actions educational institutions undertake in an effort to comply with Title IX in fact violate Title IX and equal protection principles by impermissibly discriminating against men. Thus, in trying to avoid one type of lawsuit—a claim of a violation of Title IX because of the elimination of women’s teams—educational institutions instead are finding themselves in another uncertain legal position: that of defending the elimination of men’s teams to achieve compliance.

The practice of cutting teams has become prevalent because intercollegiate athletic departments are facing ever-increasing financial pressure to reduce budgets. In making the required cuts, however, institutions run the risk of violating Title IX if such cuts reduce the opportunities for women at an institution that does not maintain substantial proportionality between its female student-athletes and its female student population. The *Cohen* court’s initial judicial and statutory interpretation has resulted in a situation where an institution that must impose budget cuts by cutting teams must comply with

---

108 Id.
109 Id.
110 See *Cohen I*, 991 F.2d at 895 (giving the interpretation *Chevron*-style deference). See infra notes 189–206 for a discussion of how the Supreme Court’s ruling in *Christensen* might have altered the *Cohen* court’s analysis.
112 See *Miami Univ. Wrestling Club*, 302 F.3d at 610; *Chalenor*, 291 F.3d at 1043; *Boulahanis*, 198 F.3d at 635; *Kelley*, 35 F.2d at 267.
113 See, e.g., *Miami Univ. Wrestling Club*, 302 F.3d at 611; *Neal* v. * Bd. of Trs.*, 198 F.3d 763, 765 (9th Cir. 1999); *Kelley*, 35 F.2d at 269.
114 See *Kelley*, 35 F.3d at 269–70; *Cohen I*, 991 F.2d 888, 898 (1st Cir. 1993).
prong one or be deemed in violation of Title IX.\textsuperscript{115} As a result of this difficulty in compliance, female athletes largely have been successful in suing institutions when their programs are cut and getting their respective athletic programs reinstated.\textsuperscript{116} This trend in case law and Title IX interpretation, coupled with the intense pressures to reduce athletic budgets, has caused institutions to resort to cutting only men’s teams in order to remain compliant with Title IX.\textsuperscript{117}

Thus far, the courts have approved these cuts as being consistent with the mandate of Title IX.\textsuperscript{118} The courts have held that a university may bring itself into compliance with Title IX either by increasing opportunities for the underrepresented gender or by decreasing athletic opportunities for the overrepresented gender.\textsuperscript{119} In coming to

\textsuperscript{115} See Cohen I, 991 F.2d at 895–900. A school complies with prong two of the three-part test if it “can show a history and continuing practice of program expansion which is demonstrably responsive to the developing interest and abilities of [the underrepresented] sex.” Policy Interpretation, supra note 43, at 71,418. Thus, an institution that cuts women’s teams will have great difficulty in complying with prong two. See Connolly & Adelman, supra note 30, at 874. Similarly, an institution that makes cuts to women’s teams will also fail prong three because courts reason that if a women’s team has been cut, the interests of those team members are not being met, which means the interests of all female athletes as a whole are not being “fully” met. See Cohen I, 991 F.2d at 898–900. Thus, schools faced with necessary budget cuts are forced to comply with prong one, substantial proportionality, or cut only men’s teams. See Connolly & Adelman, supra note 30, at 846.

\textsuperscript{116} See Roberts v. Colo. State Bd. of Agric., 998 F.2d 824, 830, 835 (10th Cir. 1993); Cohen I, 991 F.2d at 895–900, 907.

\textsuperscript{117} E.g., Miami Univ. Wrestling Club, 302 F.3d at 611 (university eliminated men’s soccer, tennis, and wrestling teams after concluding it was not in compliance with Title IX and it did not have funds to increase women’s opportunities); Boulahanis, 198 F.3d at 635 (university eliminated men’s soccer and wrestling teams after recognizing it was not in compliance with Title IX and was unable to meet the interests and abilities of females); Kelley, 35 F.3d at 269 (university eliminated men’s swimming in light of budget constraints and its failure to comply with Title IX).

\textsuperscript{118} Chalenor, 291 F.3d at 1043, 1048 (approving the elimination of men’s wrestling); Kelley, 35 F.3d at 272–73 (approving the elimination of men’s swimming while retaining women’s swimming); see Miami Univ. Wrestling Club, 302 F.3d at 611, 615–16 (approving the elimination of men’s soccer, tennis, and wrestling).

\textsuperscript{119} Neal, 198 F.3d at 771 (“[A]n institution . . . can bring itself into Title IX compliance by reducing sufficiently the number of roster spots available to men.”); Boulahanis, 198 F.3d at 638 (“[T]he elimination of men’s athletic programs is not a violation of Title IX as long as men’s participation in athletics continues to be ‘substantially proportionate’ to their enrollment.”); Horner v. Ky. High Sch. Athletic Ass’n, 43 F.3d 265, 275 (6th Cir. 1994) (“An institution need not pour ever-increasing sums into its athletic programs in order to bring itself into compliance, but has the option of reducing opportunities for the overrepresented gender while keeping opportunities for the underrepresented gender stable.”); Roberts, 998 F.2d at 830 (“Financially strapped institutions may still comply with Title IX by cutting athletic programs such that men’s and women’s athletic participation rates become substantially proportionate to their representation in the undergraduate population.”); Cohen I, 991 F.2d at 898–99 n.15 (“[A university] can . . . bring itself into
this conclusion, each court again relied upon the analysis of the First Circuit in \textit{Cohen I} and its interpretation of the Regulation and Policy Interpretation.\footnote{120}{See Neal, 198 F.3d at 770–72; Boulahanis, 198 F.3d at 639; Horner, 43 F.3d at 275; Roberts, 998 F.2d at 830 & n.9.}

\section{The First Claim: Cutting Men’s Teams Is a Violation of Title IX}

In an attempt to fight the elimination of their respective programs, male athletes first have attempted to claim that the elimination of their teams, motivated by a concern for achieving gender parity under Title IX, is itself a clear example of sex discrimination, which Title IX and its implementing regulations expressly forbid.\footnote{121}{See Miami Univ. Wrestling Club, 302 F.3d at 610; Chalenor, 291 F.3d at 1043; Boulahanis, 198 F.3d at 635.} In particular, male athletes argue that by eliminating their programs because of the sex of the participants, the institutions are discriminating against the male athletes on the basis of their sex, excluding them from participation in educational programs because of their sex, and denying them the benefits of educational programs because of their sex, all in violation of Title IX.\footnote{122}{See Miami Univ. Wrestling Club, 302 F.3d at 610.} Additionally, male athletes argue that allowing universities to make such cuts on the basis of sex, in an attempt to equalize the rates of participation and allocation of resources, is analogous to implementing a quota system based on sex.\footnote{123}{See id.} Such a quota system, they argue, is also contrary to Title IX’s purpose of encouraging, rather than reducing, athletic opportunities for all students.\footnote{124}{See id.}

Thus far, these arguments by male athletes have never succeeded in court.\footnote{125}{See Miami Univ. Wrestling Club, 302 F.3d at 610–16; Kelley, 35 F.3d at 272–73.} The U.S. Court of Appeals for the Seventh Circuit heard the first of these male athlete challenges to Title IX in 1994 in \textit{Kelley v. Board of Trustees}.ootnote{126}{35 F.3d at 267. The court was presented with an appeal from a district court’s granting of summary judgment in favor of the defendants. \textit{Id.}} In that case, the University of Illinois chose to drop its men’s swimming team in response to budget constraints and in light of the university’s failure to comply with Title IX.\footnote{127}{Id. at 269. In 1982, the DOE’s OCR determined that the University of Illinois was not in compliance with Title IX, but it granted the institution time to remedy the situation. \textit{Id.}} In re-compliance [with Title IX] . . . by reducing opportunities for the overrepresented gender while keeping opportunities stable for the underrepresented gender (or reducing them to a much lesser extent).
sponse to this action, male swimmers at the institution filed a complaint alleging that the school had violated Title IX by choosing to cut the men’s swimming program. 128

In holding that the institution had not violated Title IX by cutting the men’s program, the circuit court agreed with the Cohen court both that the Regulation and the Policy Interpretation were a reasonable interpretation of the statute and that deference should be accorded to the agency’s interpretations. 129 Thus, the court concluded, the institution did not violate Title IX when it acted in accordance with the Regulation and the Policy Interpretation. 130 The court also made clear that if the institution had cut women’s programs as part of the cost-cutting measure, the action would have violated Title IX. 131 Finally, in response to the argument that the proportionality test operates as an impermissible quota, the court held that the three-part test offered a flexible means for the university to achieve compliance, and thus a school is not forced to meet the substantial proportionality requirement. 132

The U.S. Court of Appeals for the Sixth Circuit reached a similar result in 2002 in Miami University Wrestling Club v. Miami University. 133 In 1994, in light of disparities in program offerings for men and women and the lack of additional funds to increase opportunities for female students, Miami University determined that it had to eliminate some male athletic opportunities to achieve compliance with Title IX. 134 It thus chose to eliminate the men’s soccer, tennis, and wrestling teams. 135 Male athletes filed suit against Miami University, arguing the decision to cut their teams violated Title IX. 136

The Sixth Circuit rejected the male athletes’ Title IX claims. 137 The court held that the Policy Interpretation merited substantial def-

---

128 Id. at 267.
129 Id. at 270.
130 Id. at 271–72.
131 See Kelley, 35 F.3d at 269–70.
132 See id. at 271.
133 302 F.3d at 615–16.
134 Id. at 611.
135 Id.
136 Id. at 609–10.
137 Id. at 615–16.
B. The Second Claim: Cutting Men’s Teams Is a Violation of Equal Protection

In addition to the Title IX claim, the second of the male athletes’ key claims is that an institution’s action in eliminating men’s but not women’s teams creates an illegal gender classification that violates the Equal Protection Clause of the Fourteenth Amendment. To violate the Equal Protection Clause, a state actor—such as a public university—must intentionally classify similarly situated individuals for different treatment based on some impermissible characteristic, such as race, national origin, or gender. Gender classifications are subject to an “intermediate” standard of scrutiny, which requires any discrimination based on sex to be substantially related to an important government objective. Thus, under this test, the court must determine, first, whether the government’s objective is legitimate and important, and, second, whether a direct, substantial relationship exists between the objective and the gender-based means of achieving it. In limited circumstances, a gender-based classification favoring one sex can be justified, but only if it intentionally and directly assists members of the underrepresented sex. Male athletes frequently argue that the elimination of men’s teams because of their gender does not satisfy this intermediate scrutiny standard because such actions are not substantially related to the purported goal of Title IX.

The most recent, and thus far the strongest equal protection challenge to Title IX’s regulatory framework, came in 2003 when the United States District Court for the District of Columbia heard Na-

---

138 See Miami Univ. Wrestling Club, 302 F.3d at 615. The court rejected the plaintiff’s contention that the Policy Interpretation was controlled by the less deferential Christensen standard rather than the Chevron standard. See id. (citing Christensen, 529 U.S. at 587; Chevron, 467 U.S. at 844). In particular, the Sixth Circuit thought that the Policy Interpretation was far different from the opinion letter involved in Christensen, and that because neither the regulations nor the Policy Interpretation were unreasonable, arbitrary, capricious, or manifestly contrary to Title IX, deference was warranted. See id.
139 See id. at 615 (citing Neal, 198 F.3d at 770; Cohen II, 101 F.3d 155, 174 (1st Cir. 1996)).
140 See id. at 610; Boulahanis, 198 F.3d at 639; Kelley, 35 F.3d at 272.
144 See id. at 728.
145 See Kelley, 35 F.3d at 272.
tional Wrestling Coaches Ass’n v. U.S. Department of Education. In this case, the plaintiffs sought a declaratory judgment and injunctive relief to prevent the DOE from enforcing Title IX in a manner that discriminates against male athletes. In particular, the plaintiffs directly challenged the Policy Interpretation and Policy Clarification, claiming that these policy statements force schools to eliminate men’s teams, artificially limit the size of men’s teams, and otherwise discriminate against male athletes based on sex, thereby denying them equal protection under the law.

Despite the district court’s ultimate holding that the plaintiffs lacked standing, the court did suggest in its opinion that a challenge like this one may have merit. Throughout the court’s lengthy discussion of the legislative history and judicial interpretation of Title IX and its regulatory framework, the court acknowledged that the policy statements may not have been universally accepted as reasonable interpretations of the statute and the Regulation. Further, the court emphasized that the holding that these plaintiffs lacked standing did not mean that the DOE actions were beyond the reach of judicial review and that other potential avenues existed by which individuals or institutions could challenge the regulatory framework. Thus, de-

---

146 263 F. Supp. 2d at 85. This challenge represented the first direct challenge of the validity of the Policy Interpretation and the Policy Clarification by way of an action brought against the DOE. Id. at 97.

147 Id. at 85.

148 Id.

149 Id. at 124–25. The court held that the plaintiffs lacked standing because they failed to meet the burden of pleading causation and redressability. Id. at 111. With respect to causation, the court held that the plaintiffs failed to allege that the three-part test represented a “substantial factor” in the institutions’ decision making. Id. In reaching this conclusion, the court emphasized that flexibility was key to the operation of Title IX and that it could not be shown that the three-part test and the 1996 Clarification so controlled the conduct of the institutions so as to confer standing on the plaintiffs. Id. at 114. With respect to redressability, the court held that it could not be shown that any direct action by the DOE in making new rules would force educational institutions to redress the plaintiff’s alleged injuries. Id. at 115.

150 See id. at 87–97. The court noted that the 1996 Clarification was drafted in response to concerns by members of Congress that schools were cutting men’s teams to reach substantial proportionality. Id. at 92. Additionally, the court addressed the fact that, prior to issuing the 1996 Clarification, the DOE received comments suggesting that the proposed Clarification and the three-part test it addressed were seriously flawed. Id.

151 See Nat’l Wrestling Coaches Ass’n, 263 F. Supp. 2d at 125. For example, the court stated that regulated agencies, such as the plaintiff’s educational institution members, have standing to challenge the DOE’s Regulation and subsequent interpretations. Id. Furthermore, an institution denied funding as a result of Title IX violation could challenge the agency’s authority, under Title IX, to adopt the policy interpretations at issue here. Id. Individual plaintiffs additionally have an implied right of action under Title IX to chal-
spite dismissing the case based on lack of standing, the court not only suggested that a successful direct challenge to Title IX’s regulatory framework was possible, but also provided a glimpse into what such a challenge might look like.152

III. Relation to Affirmative Action

In this most recent male challenge to Title IX, National Wrestling Coaches Ass’n v. U.S. Department of Education, the United States District Court for the District of Columbia suggested that a constitutional challenge would be possible with the right group of plaintiffs and a correctly stated claim.153 The basis of such a claim would be that the effect of the Policy Interpretation and Clarification is to cause Title IX to act as an impermissibly inflexible quota system, because there exists, in effect, only one way to comply.154 Because the crux of the argument is that Title IX is operating as an inflexible quota system, recent affirmative action caselaw provides instruction on the current U.S. Supreme Court sentiment on the use of quotas in education.155 The major Supreme Court decisions in this area are Regents of the University of California v. Bakke and the two recent decisions of Grutter v. Bollinger and Gratz v. Bollinger.156

challenge educational institutions’ conduct under the Regulation (in cutting men’s teams), and they are free to argue in such lawsuits that the statute, its Regulation, the Policy Interpretation, and/or the Policy Clarification are not entitled to deference and should be struck down as unconstitutional. Id. In stating this final suggestion, the court clarified that the holding of the Sixth Circuit in Miami University Wrestling Club did not preclude this sort of action, but rather it required that plaintiffs directly challenge the Regulation and Policy Interpretation, by way of naming the DOE as a defendant, when they challenge the conduct of the institution taken in conformity with those agency policies. Id. (citing Miami Univ. Wrestling Club, 302 F.3d at 614).

152 See id. The Court of Appeals for the District of Columbia affirmed this decision that the plaintiffs lacked standing. Nat’l Wrestling Coaches Ass’n v. Dep’t of Educ., 366 F.3d 930, 933 (D.C. Cir. 2004), cert. denied, 125 S. Ct. 2537 (2005). In affirming the decision, the court did not elaborate further on the district court’s comments about the potential that Title IX had been misinterpreted or misconstrued. See id.


154 See id.


A. Bakke—A Historical Background

In the landmark 1978 case, *Regents of the University of California v. Bakke*, the U.S. Supreme Court addressed the use of race in public higher education admissions.\(^{157}\) The Medical School of the University of California at Davis (the “Medical School”) used a special admissions program to increase the representation of disadvantaged students in each Medical School class, whereby it reserved sixteen out of 100 seats in the class for members of certain minority groups, who were subject to a separate admissions process.\(^{158}\) The plaintiff, a white male who had applied for admission at the Medical School in both 1973 and 1974, but was denied admission in both years, argued that the special admissions program operated as a racial and ethnic quota in violation of the Equal Protection Clause of the Fourteenth Amendment.\(^{159}\)

The Supreme Court issued six separate opinions in the decision, none of which garnered the support of the majority of the court.\(^{160}\) Justice Powell’s opinion announcing the judgment, however, has served as the basis for constitutional analysis of race-conscious admissions policies.\(^{161}\) In his opinion, Justice Powell concluded that race-based policies such as the Medical School’s special admissions process must be subject to strict scrutiny under the Fourteenth Amendment.\(^{162}\) He then concluded that the attainment of a diverse student body is a constitutionally permissible goal for a higher education institution, but that the Medical School’s special admissions program did not constitute a racial classification that was necessary for the attainment of that goal.\(^{163}\)

In reaching that conclusion, Justice Powell pointed to other higher education institutions that used race as one of a number of factors considered in the admissions process.\(^{164}\) He viewed the flexibility of such a process as being key to keeping it constitutionally

\(^{157}\) *Bakke*, 438 U.S. at 274–75 (opinion of Powell, J.).

\(^{158}\) Id. at 272, 274–75. This special admissions process consisted of a separate committee that considered the applications of those who stated they wished to be considered as members of a “minority group.” Id. at 274. These applications were rated, and the special committee would recommend the special applicants until the faculty-prescribed number of seats for minority students were filled. Id. at 275.

\(^{159}\) Id. at 276, 278. In 1973, at the time when the plaintiff was rejected, four special admissions slots remained unfilled, but the plaintiff was not considered for any of them. Id. at 276.

\(^{160}\) *Grutter*, 539 U.S. at 322–23 (discussing the procedural outcome of the *Bakke* decision).

\(^{161}\) Id. at 323.

\(^{162}\) See *Bakke*, 438 U.S. at 299 (opinion of Powell, J.).

\(^{163}\) See id. at 311–12, 319–20.

\(^{164}\) Id. at 316–18.
valid. Justice Powell concluded that, though race can be considered as a factor in the admissions process, the practice of setting aside a specific percentage of seats in an entering class for minorities—which in effect tells applicants who are not minorities that they are totally excluded from those seats—disregards the applicants’ individual rights guaranteed by the Fourteenth Amendment.

B. Grutter and Gratz: Current Supreme Court Thinking

Twenty-five years passed after the splintered opinion of Bakke, during which time courts and institutions struggled to determine whether Justice Powell’s opinion was binding.\textsuperscript{167} Then, in the 2003 case Grutter v. Bollinger, the U.S. Supreme Court upheld as constitutional the University of Michigan Law School (the “Law School”) admissions policy, which required admissions officials to look beyond grades and test scores to other “soft” variables that included, among other factors, race and ethnic background.\textsuperscript{168} In so doing, the Court expressly endorsed Justice Powell’s view in Bakke that diversity in the student population is a compelling state interest that justifies the narrowly tailored use of race classifications in higher education admissions.\textsuperscript{169}

Key to the Court’s determination that the Law School’s admissions policy was narrowly tailored and thus constitutional was its flexibility, which meant that race was considered in a non-mechanical way.\textsuperscript{170} The Court reiterated that to comply with the equal protection mandate of the Fourteenth Amendment, universities cannot establish quotas for members of certain racial groups or put members of those groups on separate admissions tracks.\textsuperscript{171} Rather, a higher education institution can consider race or ethnicity only as a “plus” factor, such that the institution’s admission program remains flexible enough to ensure race or ethnicity is not the defining feature of students’ applications.\textsuperscript{172} Thus, as per this recent Supreme Court decision regarding the use of race in higher education admissions, the reigning senti-

\begin{itemize}
\item\textsuperscript{165} See id. at 317–18.
\item\textsuperscript{166} Id. at 319–20.
\item\textsuperscript{167} Grutter, 539 U.S. at 322–26.
\item\textsuperscript{168} Id. at 315–16, 343.
\item\textsuperscript{169} Id. at 325.
\item\textsuperscript{170} Id. at 334.
\item\textsuperscript{171} See id. The Court defined “quota” as “a program in which a certain fixed number or proportion of opportunities are ‘reserved exclusively for certain minority groups.’” Id. at 335 (quoting Richmond v. J.A. Croson Co., 488 U.S. 469, 496 (1989)).
\item\textsuperscript{172} See Grutter, 539 U.S. at 334, 336–37.
\end{itemize}
ment emphasizes the significance of a flexible approach when using a suspect class under the Fourteenth Amendment as a means of achieving a compelling state interest.\textsuperscript{173}

The U.S. Supreme Court again emphasized the requirement of flexibility and individualized consideration in the \textit{Gratz v. Bollinger} decision, which it issued on the same day it issued \textit{Grutter}.\textsuperscript{174} In \textit{Gratz}, two white students who had been denied admission to the University of Michigan’s undergraduate program sued the institution, alleging that the admissions program violated their equal protection rights under the Fourteenth Amendment.\textsuperscript{175} The university’s admissions office used a system by which any applicant who was a member of an underrepresented racial or ethnic minority group automatically received twenty points towards the total number needed for admission.\textsuperscript{176}

In rejecting this admissions program as unconstitutional, the Court emphasized that the program was not narrowly tailored because it did not provide the individualized consideration that the Constitution required and that Justice Powell contemplated in \textit{Bakke}.\textsuperscript{177} In particular, the Court emphasized the fact that the admissions process awarded points (and subsequently admission) solely on the basis of race—without considering other factors.\textsuperscript{178} This process did not offer applicants the individualized consideration and selection process that the Equal Protection Clause required and that the Supreme Court upheld in \textit{Grutter}.\textsuperscript{179} The holdings in these affirmative action cases provide insight into current Supreme Court thinking regarding higher education decision making based on class membership—insights that become relevant in the attempt to dissect the current operation of Title IX’s regulatory framework.\textsuperscript{180}

\textsuperscript{173} See \textit{id.} at 334.
\textsuperscript{175} \textit{Gratz}, 539 U.S. at 251–52.
\textsuperscript{176} \textit{Id.} at 255.
\textsuperscript{177} \textit{Id.} at 269–70.
\textsuperscript{178} \textit{Id.} at 271–72. Indeed, the only factor that determined the distribution of points was whether the individual was a member of an underrepresented minority group. \textit{Id.}
\textsuperscript{179} Compare \textit{Gratz}, 539 U.S. at 275 (holding the policy was not sufficiently narrowly tailored to satisfy the Fourteenth Amendment), \textit{with Grutter}, 539 U.S. at 334, 337, 343 (holding that the Law School’s consideration of race as a “plus” factor retained sufficient flexibility to satisfy the Fourteenth Amendment).
\textsuperscript{180} See \textit{infra} notes 181–253 and accompanying text.
IV. Analysis: Title IX Is Operating as an Impermissibly Inflexible Quota System That Courts Must Reject as Unconstitutional

The key to unmasking the true problems with Title IX’s current regulatory framework is in examining the effects of the regulatory framework’s interpretation rather than its structure.181 Thus far, the DOE and the courts have justified the current framework, as interpreted, based on its flexibility and its importance as a governmental objective.182 Proper analysis of this framework, however, should look to how it operates in practice.183 Such an analysis would reveal that the framework, in fact, operates as an inflexible quota system that forces education institutions to cut men’s teams if they are to remain in compliance with Title IX.184 Such an inflexible system is of the type that the U.S. Supreme Court expressly rejected in its relevant affirmative action decisions, and as such, courts should reject it here as a violation of the Equal Protection Clause of the Fourteenth Amendment.185

A. The Flawed Analysis of Cohen

The problems with the current Title IX regime stem from the 1993 First Circuit Court of Appeals’ initial flawed interpretation in Cohen I.186 This case was the first challenge of an institution’s athletic programs under Title IX, and it has served as the key decision: virtually every federal circuit court of appeals subsequently facing a similar Title IX challenge has followed the First Circuit’s analysis of the Regulation and Policy Interpretation.187 The First Circuit’s analysis was flawed, however, because (1) it inappropriately granted substantial deference to the Policy Interpretation, (2) it improperly interpreted

---

181 See supra notes 113–17 and accompanying text.
183 See supra notes 113–17; 170–80 and accompanying text.
184 See Miami Univ. Wrestling Club v. Miami Univ., 302 F.3d 608, 611 (6th Cir. 2002); Chalenor v. Univ. of N.D., 291 F.3d 1042, 1043 (8th Cir. 2002); Kelley, 35 F.3d at 269.
186 See Cohen I, 991 F.2d 888, 895–900 (1st Cir. 1993); Connolly & Adelman, supra note 30, at 882.
prong three of the three-part test, and (3) it misinterpreted the proper application of the three-part test as a whole.  

1. Flaw #1: Improper Deference to the Policy Interpretation

The First Circuit’s first error was its decision to grant substantial deference to the Policy Interpretation. In making this decision, the court relied on *Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc.*, which stands for the principle that courts give great deference to an agency’s interpretation of a statute that the agency is responsible for administering. Under *Chevron*, however, regulations and their interpretations should not be given controlling weight if they are “arbitrary, capricious, or manifestly contrary to the statute.” Indeed, *Chevron* indicates that courts should give deference to such interpretations only if they reflect an agency’s reasonable construction of the statute at issue. In addition to this general caution regarding the granting of deference, the U.S. Supreme Court held more recently in 2000, in *Christensen v. Harris County*, that substantial deference should not be accorded to agency interpretations that do not have the force of law, such as those in opinion letters, policy statements, agency manuals, and enforcement guidelines. In light of the cautions under *Chevron* and the more recent ruling in *Christensen* (which the Supreme Court issued four years after the *Cohen* court ruled), strong arguments exist for questioning the appreciable deference that the First Circuit—and every subsequent circuit court—granted to the Policy Interpretation and its three-part test.

The First Circuit should not have granted substantial deference to the Policy Interpretation primarily because of the Policy Interpretation’s inconsistency with Title IX and the Regulation. First, it is unclear why the DOE, in its Policy Interpretation, chose to emphasize the first factor of the ten-factor test over the other nine. Indeed, the agency pulled out the first factor—accommodating the interests and

---

188 See infra notes 189–225 and accompanying text.
189 See Cohen I, 991 F.2d at 895.
191 Chevron, 467 U.S. at 844.
192 Chalener, 291 F.3d at 1046.
193 See id. at 1046 (citing Christensen v. Harris County, 529 U.S. 576, 587 (2000)).
194 See infra notes 195–206 and accompanying text.
195 See infra notes 196–206 and accompanying text.
abilities of male and female students—and then created a separate three-part test to evaluate compliance with that factor. In the Regulation, however, the DOE set forth the list of ten factors to serve as guidance in terms of determining compliance with the statute; nothing in the Regulation indicated that institutions, agencies, or courts should give such substantial weight to the first factor.

The Policy Interpretation is also inconsistent with Title IX itself in its use of substantial proportionality as a means of compliance. Prong one of the three-part test states that compliance with this prong can be achieved if an institution provides athletic opportunities for male and female students in proportion to their general respective enrollments in the institution. This use of statistical proportions, however, is expressly rejected as a means of finding discrimination under § 1681(b) of the statute. Furthermore, this use of the general student population for comparison purposes is too simplistic and is inconsistent with other current discrimination analyses.

Indeed, several district courts have called into question the deference afforded the Policy Interpretation with respect to the substantial proportionality prong. The U.S. District Court for the Middle District of Louisiana held in its 1996 decision, Pederson v. Louisiana State University, that the substantial proportionality prong should not be considered a “safe harbor,” because to stop the inquiry at the point of numerical equality would be inconsistent with the mandate of the statute. Similarly, the U.S. District Court for the Eastern District of

---

197 Id. at 71,414, 71,417–18.
198 See Nondiscrimination on the Basis of Sex, supra note 38, at 24,134, 24,143.
199 See infra notes 200–06 and accompanying text.
200 Policy Interpretation, supra note 43, at 71,418.
201 See supra note 34 and accompanying text.
202 See Connolly & Adelman, supra note 30, at 862–63. Indeed, it is exactly this sort of comparison to the general population that the Supreme Court has expressly rejected as being irrelevant in Title VII employment discrimination cases. See, e.g., Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 308 (1977). Rather, the Court has held that the appropriate comparison group is that of skilled or qualified applicants in the general population. Id. The appropriate comparison group in the Title IX athletics context, then, should be the population of skilled and qualified students in the general population, and the failure of the agency to use this as the proxy group for the first test renders this prong questionable at best. See Pederson v. La. State Univ., 912 F. Supp. 892, 913–14 (M.D. La. 1996) (“Pederson District”), rev’d, 213 F.3d 858 (5th Cir. 2000).
California held in its 1999 decision, *Neal v. Board of Trustees*, that relying on proportionality to achieve compliance constituted an implementation of a quota based on gender, which was a violation of Title IX.205 Thus, the fact that courts have expressed concern over the Policy Interpretation, coupled with the inconsistencies between the Policy Interpretation and the intent of the Regulation, calls into question the granting of substantial deference to the Policy Interpretation.206

2. Flaw #2: Improper Interpretation of Prong Three

The granting of deference was not the only error that the Cohen court made, however.207 The First Circuit’s second error was its interpretation of the third prong of the test—full and effective accommodation of the interests and abilities of both sexes.208 The court rejected Brown University’s contention that an institution can comply with the third prong by meeting the comparative or “relative” levels of interest of each sex—that is, by meeting the interests and abilities of both genders in an equal proportion.209 Rather, the court interpreted the standard to require “full and effective accommodation,” meaning that if sufficient interest and ability exists among members of the underrepresented gender that existing programs do not satisfy, then an institution necessarily will fail prong three of the test.210

The court justified the use of this more demanding standard, in part, because of the perceived difficulty in assessing and quantifying relative interests and abilities.211 The standard the court adopted required a simple inquiry: whether the unmet need of the underrepresented gender is sufficient to require the creation or upgrading of a team.212 Brown’s proposed standard, by comparison, would have required inquiry into what would be the appropriate survey population—the university, feeder schools, or the regional community—an inquiry the court thought would overcomplicate this already complex area of the law.213 In coming to this conclusion, however, the court

205 1999 WL 1569047, at *3.
206 See supra notes 196–205 and accompanying text.
208 See id.
209 Cohen II, 101 F.3d at 174. Under Brown’s reasoning, for example, if a school met 80% of the interests of men on a campus, then it should also meet 80% of the interests of women on campus. See id.
210 Id. (citing Cohen I, 991 F.2d at 898–99).
211 Cohen I, 991 F.2d at 900.
212 Id.
213 Id.
failed to recognize that the agency, in its Policy Interpretation, had intended institutions and courts to make this sort of relative interest inquiry.\footnote{See id.; Policy Interpretation, supra note 43, at 71,417.} Indeed, the Policy Interpretation actually provided guidance regarding methods an institution could use to assess the athletic interests and abilities of its students.\footnote{Policy Interpretation, supra note 43, at 71,417. The DOE actually provides an investigator’s manual to assist with assessing compliance; this manual also provides methods for ascertaining levels of interest, suggesting the use of surveys and other techniques. See Connolly & Adelman, supra note 30, at 865–66.} Therefore, the First Circuit’s interpretation of the third prong, expressly adopted because of its simplicity, seems to contradict the agency’s intent as reflected in the Policy Interpretation.\footnote{See Connolly & Adelman, supra note 30, at 864–67.}

3. Flaw #3: Improper Interpretation of the Three-Part Test

The Cohen court’s final error was in its interpretation of the three-part test as a whole.\footnote{See infra notes 218–23 and accompanying text.} After engaging in an overview of the Title IX Regulation and Policy Interpretation, the court then engaged in a further analysis of how these agency documents should work in practice.\footnote{See Cohen I, 991 F.2d at 895–97.} First, with respect to the first prong of the test, the court held that evidence of statistical inequality could be used as evidence of discrimination.\footnote{Id. at 895.} The court reached this conclusion despite explicit language in Title IX stating that it does not mandate statistical equality between the gender ratio in an institution’s general population and the ratio in the body of student-athletes.\footnote{Id.; see 20 U.S.C. § 1681(b) (2000).} Second, the court went on to hold that even if an institution met the “financial assistance” and “athletic equivalence” standards, it nonetheless could violate Title IX: in other words, failing the controversial three-part test under the effective accommodations area of compliance necessarily would equate to noncompliance with Title IX.\footnote{See Cohen I, 991 F.2d at 897.} Not only was this ruling expressly contrary to the language of the Policy Clarification,\footnote{See supra note 53 and accompanying text. The Policy Clarification specifically stated that the requirement addressed by the three-part test was but one of many factors considered in determining compliance. Policy Clarification, supra note 51.} but it also effectively eliminated factors two through nine of the ten-
factor test from consideration during a Title IX compliance inquiry—a result seemingly inconsistent with agency intent.223

To summarize, through its “watershed” analysis in Cohen, the First Circuit effectively construed Title IX (1) to eliminate factors two through ten from the relevant inquiry, (2) to interpret the third prong of the three-part test—accommodation of the interests of the underrepresented gender—to mean that any unmet interest would result in violation, (3) to affirm the Policy Interpretation’s use of the general population as the relevant comparison group for purposes of determining substantial proportionality under prong one, and (4) to pave the way for substantial deference to a Policy Interpretation filled with inconsistencies with the original Regulation and the statute.224 This interpretation of the Title IX statute has continued to survive, and it is this interpretation, coupled with the realities of modern fiscal constraints, that has caused Title IX to act as an impermissible gender-based quota system.225

B. The Effective Elimination of Title IX’s Flexibility

In response to charges from male athletes that Title IX, as interpreted, acts as an impermissible quota, both the DOE and the courts have highlighted the inherent flexibility built into the statute for institutional compliance.226 Indeed, in defending the substantial proportionality test, both the First Circuit and the DOE in its 1996 Policy Clarification reiterated that this was but one way to determine whether an institution’s athletics programs comply with Title IX.227 The proportionality test is considered a starting point, and as the First Circuit pointed out, the second and third parts of the test are provided in recognition of the fact that under certain circumstances, something short of this proportionality is satisfactory for determining gender balance.228 These arguments in defense of the three-part test, however, miss the point because the inflexibility has resulted not from

223 See Cohen I, 991 F.2d at 897. In the Policy Interpretation, the agency highlights factors two through ten as factors the DOE should consider in determining whether an institution is providing equal opportunity in intercollegiate athletics. Policy Interpretation, supra note 43, at 71,415.
224 See supra notes 186–223 and accompanying text.
225 See supra notes 113–17 and accompanying text.
226 Nat’l Wrestling Coaches Ass’n, 263 F. Supp. 2d at 113; Dear Colleague Letter, supra note 51; see Cohen II, 101 F.3d at 170–71; Kelley, 35 F.3d at 271.
227 Cohen II, 101 F.3d at 171; Policy Clarification, supra note 51.
228 Cohen I, 991 F.2d at 897–98.
the structure of the three-part test, but, rather, from the application of the test. 229 Indeed, the initial interpretation of the test, coupled with the financial pressures currently confronting educational institutions, has resulted in a highly inflexible test that leaves institutions with no choice but to cut men’s teams. 230

Though they espouse the built-in flexibility of the test, the courts also recognize that any effort to cut women’s sports teams would leave an institution in danger of noncompliance with Title IX. 231 The second prong of the three-part test requires institutions to have a history and continuing practice of increasing athletic opportunities for women. 232 Thus, almost by definition, an institution that cuts women’s teams necessarily will fail prong two of the test. 233 Further, as a result of the Cohen court’s interpretation of prong three, if an institution cuts women’s teams, then, unmet interest in the underrepresented gender is easily demonstrable, and thus the institution will fail prong three of the test. 234 The result is that an institution facing budget constraints is left with only prong one—substantial proportionality—as a means of complying with Title IX. 235 Yet this standard is difficult, if not impossible, to meet. 236 So, institutions then turn back to prongs two or three, which leaves them with only one option for achieving the needed budget cuts while remaining compliant with Title IX: the elimination of men’s teams. 237

Thus, this cycle reveals that the structural flexibility of the Title IX statute essentially has been interpreted away, forcing institutions either to meet statistical equality with their general populations or to eliminate only men’s teams when faced with budget cuts. 238 This inflexible system means institutions eliminate men’s teams solely because of gender and thus operates, in effect, as the type of discrimina-

---

229 See supra notes 113–17 and accompanying text.
230 See supra notes 113–17 and accompanying text.
231 Kelley, 35 F.3d at 269; see Cohen II, 101 F.3d at 171, 174–75, 180; Cohen I, 991 F.2d at 898–900.
233 Connolly & Adelman, supra note 30, at 874.
235 See id.
236 See Connolly & Adelman, supra note 30, at 862–64.
237 See Miami Univ. Wrestling Club, 302 F.3d at 611; Chalenor, 291 F.3d at 1043; Kelley, 35 F.3d at 269.
238 See Miami Univ. Wrestling Club, 302 F.3d at 611; Chalenor, 291 F.3d at 1043; Kelley, 35 F.3d at 269.
tory quota system that the Supreme Court has expressly rejected in its recent affirmative action decisions. 239

C. An Inflexible Quota Is Unconstitutional Under the U.S. Supreme Court’s Affirmative Action Cases

The U.S. Supreme Court’s recent rulings on affirmative action in higher education admissions provide insight to the Title IX discussion. 240 First, in these cases, the Court has held that student body diversity is a compelling state interest and thus that the narrowly tailored use of race to achieve this compelling state interest is justified. 241 Second, and more importantly, an institution that wishes to use race as a criterion in its decision-making process must do so in a “flexible, non-mechanical” way; only this flexibility allows such a policy to remain narrowly tailored enough to be constitutional. 242 Indeed, in upholding the Law School’s admission policy in Grutter v. Bollinger, the Court emphasized the flexibility of the Law School’s approach, in which race and ethnic background were considered a “plus” factor in the admissions process, but each decision remained highly individual. 243 Likewise, the Court emphasized the non-flexibility of the schools’ approaches when it rejected the admissions policies used in both Regents of the University of California v. Bakke and Gratz v. Bollinger. 244 In Bakke, the institution used a racial set-aside program whereby a certain number of seats were reserved for members of minority groups; in Gratz, the institution automatically granted extra points that counted toward automatic admission to applicants of particular racial and ethnic backgrounds. 245 In both decisions, the Court emphasized that such broad, inflexible standards were not narrowly tailored, and as such the institutions had violated the Equal Protection Clause of the Fourteenth Amendment. 246

The Court’s analysis in each of these three cases is instructive as to how courts should view the current operation of the Title IX statute. 247 Title IX’s operation is analogous to that of affirmative action

240 See infra notes 241–46 and accompanying text.
241 Grutter, 539 U.S. at 325.
242 See id. at 334.
243 Id. at 315–16, 334.
244 See Gratz, 539 U.S. at 270–72; Bakke, 438 U.S. at 319–20 (opinion of Powell, J.).
245 Gratz, 539 U.S. at 270; Bakke, 438 U.S. at 274–75 (opinion of Powell, J.).
246 See Gratz, 539 U.S. at 270–72; Bakke, 438 U.S. at 319–20 (opinion of Powell, J.).
247 See infra notes 248–53 and accompanying text.
programs in that, as currently interpreted, it forces universities to make decisions based on an individual's membership in a particular group—gender.\(^\text{248}\) Thus, it should follow that the key to withstanding constitutional scrutiny of such decisions is to emphasize the flexibility of the system, which can demonstrate that individual rights are not being curtailed.\(^\text{249}\) As discussed above, however, any flexibility that had been built into the Title IX statute now, in effect, has been interpreted out of it, such that an institution facing budget constraints has no choice but to cut men’s teams or be in violation of Title IX.\(^\text{250}\) Though gender-based decisions are subject to the lesser standard of intermediate scrutiny (as opposed to the strict scrutiny of race-based decisions), it does not follow that the emphasis on flexibility in the process should be correspondingly lessened.\(^\text{251}\) The appropriate degree of flexibility remains to be determined, but the current lack of any flexibility at all is certainly not preferable.\(^\text{252}\) The decisions being made under Title IX to cut men’s teams would not pass constitutional muster if they were based on race, and the current scheme also should not pass constitutional muster when the decisions are based on gender.\(^\text{253}\)

D. The Future of Title IX: The Need for Federal Circuit Court Action

Guided by the U.S. Supreme Court’s recent affirmative action holdings in Grutter and Gratz, informed by the Court’s 2000 Christensen decision, and enlightened by the substance of the recent male athletic challenges to the Title IX statute, the federal circuit courts of appeals should alter their current analysis of Title IX claims and thereby prevent Title IX from continuing down its discriminatory path.\(^\text{254}\) Although the First Circuit’s 1993 Cohen I decision was a noble attempt at defining an appropriate enforcement structure, it was not without flaws—flaws that recent financial struggles in higher education have

---

\(^{248}\) See Miami Univ. Wrestling Club, 302 F.3d at 611; Chalenor, 291 F.3d at 1043; Kelley, 35 F.3d at 269.

\(^{249}\) See Grutter, 539 U.S. at 336–37; Gratz, 539 U.S. at 270–72; Bakke, 438 U.S. at 319–20 (opinion of Powell, J.).

\(^{250}\) See Miami Univ. Wrestling Club, 302 F.3d at 611; Chalenor, 291 F.3d at 1043; Kelley, 35 F.3d at 269.


\(^{252}\) See supra notes 226–39 and accompanying text.

\(^{253}\) See Gratz, 539 U.S. at 270–72; Bakke, 438 U.S. at 319–20 (opinion of Powell, J.).

\(^{254}\) See infra notes 255–74 and accompanying text.
The current legal landscape and fiscal realities in higher education make this a time characterized by new issues that the Cohen court did not anticipate when it rendered its decision almost fifteen years ago. It is time for the federal circuit courts of appeals to stop according judicial deference to the Cohen court’s original interpretation and to reevaluate Title IX’s regulatory framework.

The circuit courts first must explicitly recognize that the Cohen court’s interpretation was flawed, particularly in its granting of substantial deference to the Policy Interpretation. The legal landscape with respect to deference to agency interpretations has changed since Cohen I. In particular, the recent Christensen decision—a decision the Cohen court did not have the luxury to use as guidance—would counsel against the deference that the First Circuit granted. The courts should recognize that not only does the Policy Interpretation not have the force of law and thus it is not entitled to substantial deference under Christensen, but also that careful analysis of the Policy Interpretation in comparison to the Regulation and the Title IX statute exposes inconsistencies that even the Chevron standard recognizes as reasons counseling against deference.

In addition to acknowledging this improper deference to the Policy Interpretation, the circuit courts also should reevaluate the substance of male athletes’ claims under Title IX and recognize that the Cohen court’s interpretation of the Policy Interpretation’s third prong and of the operation of its entire three-part test was flawed and has led to an impermissibly inflexible system. In engaging in this evaluation, the courts can use reasoning and analysis similar to that used by the U.S. Supreme Court in Grutter and Gratz. In those affirmative action decisions, the Court demanded that decision making based on group classifications (for example, race or gender) be flexible so as to respect the rights of individuals. The circuit courts

---

255 See supra notes 189–225 and accompanying text.
256 See supra notes 111–20 and accompanying text.
257 See supra notes 189–206 and accompanying text.
258 See supra notes 189–94 and accompanying text.
259 See Chalenor, 291 F.3d at 1046 (citing Christensen, 529 U.S. at 587).
260 See supra notes 189–206 and accompanying text.
261 See supra notes 207–25 and accompanying text.
262 See supra notes 240–53 and accompanying text.
should recognize this demand when analyzing Title IX claims. The current operation of the regulatory framework results in an inflexible system whereby universities are forced to make decisions based solely on gender. The U.S. Supreme Court’s affirmative action holdings reject such inflexible decision making, and guided by those holdings, circuit courts likewise should reject it here.

Having discovered the inflexibility of the current regulatory framework, the circuit courts must hold the current framework unconstitutional. Then, the courts should advise the DOE to reexamine and reinterpret Title IX and the Regulation so as to provide a policy interpretation and accompanying tests that allow for flexibility and full individual consideration in the decision-making process. Potential reformulations of the agency’s interpretation of the statute could call for a more balanced inquiry that reinstates examination of the other nine factors originally set forth in the Regulation. Additionally, a reexamination and perhaps reinterpretation of prongs one and three of the three-part test also would be advisable.

It is important to note that the ultimate goal of such a reexamination is not to get rid of Title IX. Few will deny the positive effects it has had for female athletes over its more than twenty-five years of existence. Rather, the DOE must revisit and revise the regulatory framework in place to enforce the statute so that it operates the way it was originally intended to operate. In that way, the DOE can prevent discrimination based on sex in a manner that does not hinder the success of any of this country’s student-athletes.

---

265 See Miami Univ. Wrestling Club, 302 F.3d at 611; Chalenor, 291 F.3d at 1043; Kelley, 35 F.3d at 269.
266 See Gratz, 539 U.S. at 270–72; Bakke, 438 U.S. at 319–20 (opinion of Powell, J.).
267 See supra notes 217–25 and accompanying text.
268 See supra notes 199–216 and accompanying text.
269 See supra notes 268–70 and accompanying text.
271 See supra notes 268–70 and accompanying text.
272 See supra notes 238–39 and accompanying text.
**Conclusion**

A close examination of the current operation of the regulatory framework for Title IX’s athletics provisions—a framework that has resulted from years of circuit court deference to the DOE’s Policy Interpretation and to the First Circuit Court of Appeals’s interpretation of that document in *Cohen v. Brown University*—reveals an impossibly inflexible system with which universities must comply. While the First Circuit’s original interpretation of and deference given to the Policy Interpretation may have been appropriate at the time, subsequent caselaw and a change in the nature of the claims under Title IX reveal the flaws in that original analysis and counsel the circuit courts to revisit the First Circuit’s decisions. In particular, the First Circuit’s holding created a system that operates in an inherently inflexible manner, in which the only means for a financially strapped institution to achieve compliance is to cut men’s athletic teams. Such an inflexible system is counter to the U.S. Supreme Court’s current thinking regarding the use of classifications in higher education decision making, as seen in the Court’s recent affirmative action holdings. With this in mind, it is time for the circuit courts to alter their analyses of the Title IX regulatory framework, recognize its inherent inflexibility as unconstitutional, and force the DOE to reinterpret Title IX and the Regulation so that Title IX can operate the way it was originally intended to operate.

Jennifer R. Capasso
Abstract: When considering applications of the Takings Clause to appropriations of interest and principal, courts rely on the quaintly expressed 18th century rule that “interest shall follow the principal, as the shadow the body.” Since that time, an elaborate body of regulatory takings law has evolved. As part of this evolution, courts delved into questions of temporary takings of property, occurring both as a result of outright physical occupation and through regulatory limitations. This Note argues that courts should reject the old “interest follows principal” rule and adopt the modern temporary takings rules. In particular, this Note explores the results of such a change in three cases involving prisoners alleging an unconstitutional taking when their interest was deposited in accounts with the Departments of Corrections while incarcerated. Adopting a temporary takings analysis would not only bring consistency between the analyses of takings of money and takings of land, but also offer a more consistent outcome within this group of prisoner interest cases.

Introduction

The law surrounding regulatory takings has been filled with contradictory holdings and conflicting justifications since its inception in 1922 with Pennsylvania Coal Co. v. Mahon.\(^1\) Meanwhile, the number and variety of situations to which the caselaw has been applied has been steadily expanding.\(^2\) One interesting expansion of the law has been in the area of takings of money, and in particular, interest.\(^3\) Cases in this line have been decided according to a traditional common law “interest follows principal” rule, which begins by asking whether or not principal is “owned” by the claimant, and if so, concludes by finding that any loss of interest is a compensable taking.\(^4\)

---

\(^1\) See 260 U.S. 393, 413 (1922); see also Jed Rubenfeld, Usings, 102 Yale L.J. 1077, 1078 n.2 (1993) (providing a synopsis of scholarly opinions on the chaos in the field).


\(^4\) See Brown, 538 U.S. at 235; Phillips, 524 U.S. at 165–66; Webb’s, 449 U.S. at 162.
The difficulty with the common law analysis is that it presumes two distinct, severed items, the principal and the interest it earns.\(^5\) Regulatory takings law, by comparison, has developed to eliminate the kind of severance this analysis represents, making that two-part inquiry appear artificial.\(^6\)

A strand of recent cases involving the interest earned on prisoners’ accounts continues to present the question of how takings of interest should be treated under regulatory takings law.\(^7\) Federal circuit courts of appeals have reached different results due to the courts’ disparate treatment of the inmates’ underlying property interest.\(^8\) This Note attempts to sidestep that disagreement by rejecting the severed principal/interest analysis and instead applies a regulatory takings approach that looks at the parcel as a whole.\(^9\) Such an approach brings consistency to the treatment of regulatory takings cases by bringing the prisoner interest cases under the same caselaw that dominates the rest of regulatory takings claims.\(^10\)

In 2004, the Eleventh Circuit Court of Appeals in *Givens v. Alabama Department of Corrections* ruled that for purposes of the Takings Clause\(^11\), inmates had no protected property interest in the interest that accrued on principal earned from work-release wages.\(^12\) This ruling permitted the Alabama Department of Corrections to retain the interest earned on inmate accounts.\(^13\) The Eleventh Circuit’s holding largely followed the reasoning laid out in 2000 in *Washlefske v. Winston*, in which the Fourth Circuit Court of Appeals found that prisoners do not have a property interest in the interest earned on principal accumulated from payments for their labor while in prison.\(^14\) In 1998, in *Schneider v. California Department of Corrections*, the Ninth Circuit Court of Appeals reached the opposite conclusion, holding that because prisoners had a property interest in the principle in their prison ac-

---

5 See Brown, 538 U.S. at 234–35; Phillips, 524 U.S. at 164; Webb’s, 449 U.S. at 160–61.
7 See Givens v. Ala. Dep’t of Corr., 381 F.3d 1064, 1068 (11th Cir. 2004); Washlefske v. Winston, 234 F.3d 179, 185 (4th Cir. 2000); Schneider v. Cal. Dep’t of Corr., 151 F.3d 1194, 1199 (9th Cir. 1998).
8 See Givens, 381 F.3d at 1068; Washlefske, 234 F.3d at 185; Schneider, 151 F.3d at 1199.
9 See infra notes 133–238 and accompanying text.
10 See infra notes 65–124 and accompanying text.
11 U.S. Const. amend. V.
12 381 F.3d at 1070.
13 See id.
14 See Washlefske, 234 F.3d at 180–81.
count, that interest extended to the interest earned.\textsuperscript{15} Under this ruling, the California Department of Corrections did not retain the interest earned on inmate accounts.\textsuperscript{16} In each of the above examples, the court looked to the common law rule that “interest follows principle.”\textsuperscript{17}

Part I of this Note discusses the history of the common law “interest follows principal” rule and applications of the rule to prisoner interest cases.\textsuperscript{18} Part II outlines a brief history of regulatory takings, the emergence of conceptual severance, and the denominator problem as formidable obstacles to developing a consistent framework for analysis.\textsuperscript{19} It characterizes the separation of ownership of interest and principal in the regulatory takings analysis as one more example of conceptual severance.\textsuperscript{20} Part III applies a regulatory takings analysis to the prisoner interest cases by first choosing the most appropriate test and then applying it to find no taking.\textsuperscript{21}

I. Evolution of the Interest Follows Principal Rule and Its Application to Takings

A. The History of the Interest Follows Principal Rule

The traditional common law rule that “interest follows principle” dates to at least the eighteenth century.\textsuperscript{22} The United States Supreme Court first applied the rule in a Takings Clause context in 1980 in \textit{Webb’s Fabulous Pharmacies v. Beckwith.}\textsuperscript{23} In \textit{Webb’s}, the Court found that for purposes of the Takings Clause, a party had a protected property interest in the interest earned on the monies deposited in an interpleader fund, which was required by state law to be deposited with the registry of the court while creditors were paid.\textsuperscript{24} The State of Florida argued that while deposited with the court, the principal and the in-

\textsuperscript{15} See 151 F.3d at 1201.
\textsuperscript{16} See id.
\textsuperscript{17} See Givens, 381 F.3d at 1067; Washlefske, 234 F.3d at 184; Schneider, 151 F.3d at 1198.
\textsuperscript{18} See infra notes 22–64 and accompanying text.
\textsuperscript{19} See infra notes 65–132 and accompanying text.
\textsuperscript{20} See infra notes 125–32 and accompanying text.
\textsuperscript{21} See infra notes 133–238 and accompanying text.
\textsuperscript{22} See Himely v. Rose, 9 U.S. (5 Cranch.) 313, 319 (1809) (holding that “interest goes with the principal as the fruit with the tree”); Beckford v. Tobin, 27 Eng. Rep. 1049, 1051 (Ch. 1749) (holding that “interest shall follow the principal, as the shadow the body”).
\textsuperscript{23} See 449 U.S. 155, 162 (1980).
\textsuperscript{24} Id. at 164–65.
terest belonged to the State as “public property.” During that time, the State argued, the depositor and the depositor’s creditors had no property interest in the principal, and so no property interest in the interest it earned. The language the Court used in describing its reasoning highlights the overlapping logic of both the “interest follows principal” rule and the more modern doctrine of temporary takings. The Court stated the following:

The earnings of a fund are incidents of ownership of the fund itself and are property just as the fund itself is property. The state statute has the practical effect of appropriating for the county the value of the use of the fund for the period in which it is held in the registry.

In its analysis the Court focused on the ownership of the principal during the relevant time period and described the ability to earn interest on that principal as incidental to the principal itself. This suggests an analysis that considers the right to earn interest as just one strand in a bundle of property rights attaching to the ownership of the principal, rather than looking at the ownership of the interest as a separate and distinct entity from the ownership of the principal. The Court also rejected the State’s argument that appropriating the interest could be characterized as a fee imposed by the court because a separate amount, limited by statute, had already been deducted from the principal for the payment of court fees.

More recently, in 1998, in Phillips v. Washington Legal Foundation, the Supreme Court applied the “interest follows principal” rule to Interest on Lawyers Trust Accounts (“IOLTA”) and held that lawyers’ clients have a protected property interest in the interest earned under IOLTA legislation. Because the funds held by lawyers were unarguably the property of the clients, the only question for the Court was whether or not the clients also “owned” the interest those funds gen-

---

25 Id. at 158–59.
26 Id.
27 See id. at 164.
28 Webb’s, 449 U.S. at 164.
29 See id.
30 See id.
31 See id. at 162.
32 524 U.S. 156, 160, 172 (1998). In 2003, in Brown v. Legal Foundation of Washington, the Court went on to find that although there was a protected property interest, no compensation was due because there was no value in the interest to the lawyers’ clients. See 538 U.S. 216, 240 (2003).
erated while held by the lawyer.\textsuperscript{33} The Court reiterated the holding of \textit{Webb's}, that a state may not appropriate the property interest created through the traditional rule that “interest follows principal” by attempting to disavow that rule with a statute.\textsuperscript{34} It noted, in language echoing \textit{Webb's}, that one of the “incidents of ownership” of principal was the ability to distribute the interest it earned.\textsuperscript{35} The Court’s language describing the ownership of the interest suggested that it was merely another aspect of the ownership of the principal, or, in other words, another stick in the bundle of rights which attaches to the principal.\textsuperscript{36} The analysis the Court actually applied, however, carefully separated the question of the ownership of the interest and the ownership of the principal.\textsuperscript{37}

Justice Stephen Breyer’s dissenting opinion in \textit{Phillips} attempts to reframe the question, transforming it from one of separate interests in principal and interest to one of a temporary intrusion into the property interests of the principal.\textsuperscript{38} This Note explores the implications of Justice Breyer’s alternative analysis by abandoning the old “interest follows principal” rule, which analyzes interest and principal separately, in favor of a temporary takings analysis that treats interest as a right of ownership that attaches to principal.\textsuperscript{39} It determines how such an alternative analysis might shed light on the different conclusions the federal circuit courts of appeals have reached on whether or not a department of corrections may appropriate the interest earned on a prisoner’s principal without compensation.\textsuperscript{40}

\textbf{B. Application of the Interest Follows Principal Rule to the Prisoner Interest Cases}

In \textit{Givens v. Alabama Department of Corrections}, a case decided by the Eleventh Circuit Court of Appeals in 2004, inmates in Alabama prisons who participated in a work-release program were required by state law to have their wages deposited with the Department of Cor-

\textsuperscript{33} See \textit{Phillips}, 524 U.S. at 164.
\textsuperscript{34} \textit{Id.} at 167.
\textsuperscript{35} See \textit{id.}
\textsuperscript{36} See \textit{id.} at 164.
\textsuperscript{37} See \textit{id.}
\textsuperscript{38} See \textit{Phillips}, 524 U.S. at 180 (Breyer, J., dissenting). Justice Breyer stopped short of the perspective this Note advocates, however, because his analysis goes on to identify a claimant’s right to earn interest independent from any other rights attaching to ownership of principal. See \textit{id.} at 181–82.
\textsuperscript{39} See \textit{infra} notes 133–238 and accompanying text.
\textsuperscript{40} See \textit{infra} notes 133–238 and accompanying text.
Fees were deducted from this account by the Department of Corrections, but state law imposed a limit on the percentage of the wages that could be deducted, limiting the Department to taking forty percent of the inmates’ wages as fees. Alabama deducted the maximum amount allowed by statute and then deposited the remainder of the wages in an account with the Department of Corrections in the prisoner’s name. Although the Department of Corrections placed the prisoners’ wages in interest-bearing accounts, the Department retained the interest earned on the wages. The court sided with the Department of Corrections, holding that there was no taking when the Department appropriated the interest. The court began its analysis by stating that the Takings Clause only protects preexisting property interests (and cannot create property interests) and such property interests must be found in state law. The court noted the common law rule that interest follows principal. It also noted, however, that at common law, prisoners had no ability to earn wages, let alone interest. Furthermore, although legislation granted inmates a property interest in the principal earned from their wages, the statute granted no corresponding property interest in the interest earned on those wages. Because such a property interest had not existed at common law, without the statutory grant, the prisoners had no claim to the interest their principal earned.

In 2000, in Washlefske v. Winston, the Fourth Circuit Court of Appeals addressed an inmate’s claims that the Virginia Department of Corrections violated the Takings Clause when it retained the interest earned on the principal in his account, called a “spend account,” during his incarceration. Unlike the Alabama statute in Givens, the Vir-

---

41 See 381 F.3d 1064, 1065 (11th Cir. 2004).
42 See id.
43 See id.
44 See id. By comparison, in Washlefske v. Winston and Schneider v. California Department of Corrections the interest was deposited into a fund for the benefit of prisoners, invoking some faint reciprocity of advantage argument for the Department. See Washlefske v. Winston, 234 F.3d 179, 181 (4th Cir. 2000); Schneider v. Cal. Dep’t of Corr., 151 F.3d 1194, 1195–96 (9th Cir. 1998). In Givens there is no mention of how the appropriated interest is spent. See 381 F.3d at 1065.
45 See Givens, 381 F.3d at 1070.
46 Id. at 1067.
47 See id.
48 Id. at 1068–69.
49 See id. at 1070.
50 Givens, 381 F.3d at 1069.
51 234 F.3d at 180. The principal in the account was earned through payments in compensation for labor while in prison, at a rate of 90 cents/hour. Id. at 181. The interest ap-
Virginia statute in *Washlefske* enabled inmates to send funds outside the prison to be invested in interest bearing accounts on their behalf. The “spend accounts” at issue in *Washlefske* were maintained for the purpose of allowing the inmates a source from which to spend money while in prison. The court began, as the Eleventh Circuit later would in *Givens*, by noting that at common law a prisoner had no property interest in the results of his labor. Any property interest an inmate had in money earned from labor, the court reasoned, must be the result of a specific grant of a property interest by state statute. In this case, the statute which granted inmates the right to earn wages also explicitly dictated that control over the interest earned from those wages would vest in the prison official, not in the inmate himself. The statute only “gave” the wages to inmates for use while in prison under certain limited circumstances, and the limited rights granted by statute not only did not mention interest with respect to inmates, but explicitly gave the ownership of the interest to the Department of Corrections.

In 1998, the Ninth Circuit Court of Appeals came to a different conclusion in *Schneider v. California Department of Corrections*, holding that the right to receive interest income generated on the principal was a sufficiently fundamental property interest associated with the ownership of the principal that a state could not appropriate the interest without violating the Fifth Amendment. The *Schneider* court acknowledged that the statute did not explicitly grant ownership of the interest itself to the inmates—the same argument later presented in *Washlefske* and *Givens*. In fact, in *Schneider*, as in *Washlefske*, a statute explicitly assigned ownership of the interest to the Department of Corrections. Nevertheless, the *Schneider* court noted that the existence of a statute giving the property interest to someone other than the inmate was not sufficient to show that the inmate had no right to the interest. Citing

propriated by the Department of Corrections was used for “the general benefit of inmates under the State’s care.” *Id.*

52 See *id.*

53 *Id.*

54 See *id.* at 184–85. The court went further, pointing out that at early common law an inmate had no right to hold personal property at all. *See id.* at 185.

55 See *id.*

56 See *Washlefske*, 234 F.3d at 185.

57 See *id.* at 181.

58 151 F.3d 1194, 1201 (9th Cir. 1998).

59 See *id.* at 1199.

60 See *id.*

61 *Id.*
Webb’s, in which a state statute entitled the court clerk to keep the interest, and Phillips, in which state law required IOLTA contributions, Schneider held that under some circumstances a state’s declaration that no property right existed was not sufficient to extinguish such a property right. In weighing the statute which eliminated the inmate’s property interest against the common law rule that “interest follows principal,” the court found the common law rule prevailed because of its long pedigree and near universal endorsement by American courts. The court held that the “interest follows principal” rule to be sufficiently fundamental that a state violated the Fifth Amendment if it attempted to change that rule by statute.

II. The Evolution of Regulatory Takings Law and Its Application to Takings of Interest

Regulatory takings analysis, compared to the common law “interest follows principal” rule, has more recent origins. For the first 130 years after the adoption of the U.S. Constitution, the protection granted in the Takings Clause of the Fifth Amendment was understood to be limited to require compensation for physical appropriations of property under the state’s power of eminent domain. In 1922, the United States Supreme Court decided Pennsylvania Coal Co. v. Mahon, which for the first time required compensation for a taking that resulted not from the physical invasion or appropriation of property, but from the imposition of government regulations that limited the owner’s use of the property. Whereas in one sense this innovation addressed the inequity of compensating for physical appropriations and not regulations imposing comparable economic harm, it also created a set of new complications. In the wake of the Pennsyl-

---

62 See id.
63 See Schneider, 151 F.3d at 1201.
64 See id. The court in Schneider comes to this conclusion with no mention of the competing common law tradition that prisoners have no rights to earn wages or interest. See id. But see Givens, 381 F.3d at 1068; Washlefske, 234 F.3d at 186.
66 See U.S. Const. amend. V; Transportation Co. v. Chicago, 99 U.S. 635, 642 (1879); Legal Tender Cases, 79 U.S. (12 Wall.) 457, 551 (1871).
vania Coal decision, courts were left to decide what regulations were legitimate exercises of the police power of the state (like zoning regulations and nuisance laws) and what regulations were so excessive that they required compensation under the Fifth Amendment. This proved to be a challenging task.

A. The Penn Central Multifactor Test, the Denominator Problem, and the Question of Conceptual Severance

The modern multifactor approach to regulatory takings analysis was outlined by the U.S. Supreme Court in 1978 in Penn Central Transportation Co. v. City of New York. In that case, the Court applied an “essentially ad hoc, factual inquiry” to determine whether or not a governmental action rose to the level of a taking. One of the factors the Court considered was the economic impact of the regulation. A second factor was the regulation’s interference with distinct, investment-backed expectations of the claimant. The third factor was the character of the governmental action, including whether or not there was a physical invasion of the property, as well as determining whether or not the regulation arose merely “from some public program adjusting the benefits and burdens of economic life to promote the common good.” Although courts have struggled with how to apply these factors, they have remained the primary framework for deciding regulatory takings cases. In combination, the three factors compare the economic impact of the regulation in question on the

---


70 See, e.g., Penn Cent. Transp. Co. v. City of New York, 438 U.S. 104, 123 (1978); see also Epstein, supra note 68, at 107; Rubenfeld, supra note 1, at 1088–91 (reviewing the conflict between various factors and principles used in defining takings).


72 Id.

73 Id.

74 Id.

75 Id.

individual property owner with the extent and importance of the public benefit it creates.\textsuperscript{77}

In 1992, the U.S. Supreme Court carved out an exception to the multifactor test articulated in \textit{Penn Central}.\textsuperscript{78} In \textit{Lucas v. South Carolina Coastal Council}, the Court adopted a per se rule that declared that in the “extraordinary circumstance” in which a regulation caused a property owner to lose all economically beneficial uses of the land, a taking necessarily resulted.\textsuperscript{79} The Court justified this rule by looking at the background principles and common law rules that predated the regulation in question.\textsuperscript{80} To support this rule, the Court noted that real property owners hold title with the expectation that it will be subject to \textit{some} regulation that might limit its value, but not that \textit{all} economic value could be lost through regulation.\textsuperscript{81} Only when preexisting background principles or common law suggested that the owner should have expected the possibility of losing all economic uses could such a regulation have avoided the requirement of compensation.\textsuperscript{82}

The introduction of the \textit{Lucas} per se rule exacerbated a problem inherent in even the very first regulatory takings cases—that of determining against which parcel the diminution in value should be measured—a problem often referred to as finding the appropriate denominator.\textsuperscript{83} Justice Brandeis raised the difficulty of determining appropriate denominators in his dissenting opinion in \textit{Pennsylvania Coal}, noting that in order to evaluate the extent of the diminution of property value, a point of reference is needed to be meaningful.\textsuperscript{84} The harm resulting from a regulation will vary depending on whether one measures the loss against just the portion of the parcel affected by the regulation, the parcel as measured by its physical boundaries, or


\textsuperscript{78} \textit{Lucas}, 505 U.S. at 1019.


\textsuperscript{80} See \textit{Lucas}, 505 U.S. at 1027.

\textsuperscript{81} See id. at 1027–28. The \textit{Lucas} rule is explicitly limited to land. See id. The Court stated that “in the case of personal property, by reason of the State’s traditionally high degree of control over commercial dealings, [an owner of personal property] ought to be aware of the possibility that new regulation might even render his property economically worthless.” See id. at 1027. This limitation lends support to the argument that the \textit{Lucas} per se rule has no place in a Takings Clause case concerning money. See id.

\textsuperscript{82} See id.

\textsuperscript{83} See id. at 1016 n.7; \textit{Pa. Coal}, 260 U.S. at 419 (Brandeis, J., dissenting).

\textsuperscript{84} See 260 U.S. at 419 (Brandeis, J., dissenting).
even the parcel as measured against all the property held by the owner affected by the regulation. Manipulating denominators along physical boundaries and conceptual lines complicates the measurement problem. Conceptually, property owners could divide ownership into the individual “sticks” of the bundle representing the collection of rights with respect to property. According to such a division, any reduction in the ability of an owner to use property might be recharacterized as a total loss of one individual stick in the bundle—the stick corresponding to the use limited or forbidden by the regulation. This manipulation is often referred to as “conceptual severance.” Such a characterization has the potential to at least strengthen the economic impact prong of the multifactor Penn Central test and to at most invoke the per se “total economic loss” rule created in Lucas in all but select cases involving relatively small losses of value.

The application of the “economic impact” and “investment-backed expectations” prongs of the multifactor test in Penn Central illustrates the effect of the denominator problem. In Penn Central, the owners of Grand Central Station claimed that they lost all value of the right to develop the space above the station when the Landmarks Preservation Commission designated Grand Central Station a landmark under New York City’s Landmarks Preservation Law, thereby

---

85 See Michelman, supra note 68, at 1192; Rose, supra note 77, at 566–69; Rubenfeld, supra note 1, at 1090–91 (discussing the results of the denominator problem as it appears in Penn Central and Keystone). Compare Pa. Coal, 250 U.S. at 414 (holding that laws restricting coal mining operations resulted in a taking), with Keystone Bituminous Coal Ass'v. DeBenedictis, 480 U.S. 470, 497–502 (1987) (finding nearly identical laws resulted in no taking, due primarily to a manipulation of the relevant denominator).

86 See, e.g., Andrus v. Allard, 444 U.S. 51, 65–66 (1979) (rejecting an attempt at severing the right to alienation); Penn Cent., 438 U.S. at 130 (rejecting an attempt at severing air rights); see also Michelman, supra note 68, at 1192–93 (describing the results of conceptual severance); Margaret Jane Radin, The Liberal Conception of Property: Cross Currents in the Jurisprudence of Takings, 88 COLUM. L. REV. 1667, 1676 (1988) (describing the emergence of conceptual severance).

87 See, e.g., Andrus, 444 U.S. at 65–66; Penn Cent., 438 U.S. at 130; see also Michelman, supra note 68, at 1193 (describing the results of conceptual severance); Radin, supra note 86 (describing an application of conceptual severance).

88 See Lucas, 505 U.S. at 1016 n.7; Pa. Coal, 260 U.S. at 419, (Brandeis, J., dissenting); see also Michelman, supra note 68, at 1193 (describing the illogical extremes of conceptual severance); Radin, supra note 86 (describing some applications of conceptual severance, and the Court’s wariness).

89 See Tahoe-Sierra, 535 U.S. at 331; Radin, supra note 86.

90 See, e.g., Tahoe-Sierra, 535 U.S. at 331; see also Michelman, supra note 68, at 1192 (describing the results of conceptually severing elements of property ownership).

91 See Penn Cent., 438 U.S. at 124–38.
preventing plans to build a fifty-story tower on top of the building.\textsuperscript{92} Even before the introduction of the \textit{Lucas} per se rule for total economic loss, Penn Central argued that such a significant deprivation should weigh favorably for determining the economic impact of the regulation on its investment-backed expectations.\textsuperscript{93} The \textit{Penn Central} Court, however, declined to look exclusively at the value of the “air rights” and instead, measured the economic impact against the company’s investment-backed expectations in the property as a whole.\textsuperscript{94}

Using the investment of a property owner to measure the economic impact of a regulation, as this method purports to do, could result in inequitable results.\textsuperscript{95} For example, compare the interference with the investment of a long-time property owner who has recovered any initial investment in the property with a recent investor who invested a similar amount but has yet to see a return on the investment.\textsuperscript{96} Those two owners, with similar investments in similar properties, might be treated differently under the Takings Clause due largely to the length of time they owned the property in question.\textsuperscript{97} Similarly, it is not clear that the economic impact factor would have resulted in the same outcome in \textit{Penn Central} if applied to a developer who had purchased the same property, only a few years before, intending to develop a skyscraper.\textsuperscript{98} As an illustration of that point, consider that of particular significance to the Court in \textit{Penn Central} was the fact that the property could still be used profitably as a train station and rental office space.\textsuperscript{99} In fact, the Court went beyond even the investment in the parcel in question in applying the economic impact factor, noting that the designation of landmark status and the transferable development rights that accompanied such status would benefit other property held by the same company.\textsuperscript{100} Essentially, the Court ex-

\textsuperscript{92} Id. at 119.
\textsuperscript{93} Id. at 130.
\textsuperscript{94} See id.
\textsuperscript{95} See infra notes 96–98 and accompanying text.
\textsuperscript{96} See Penn Cent., 438 U.S. at 135–38 (discussing length of ownership as a factor in takings analysis).
\textsuperscript{97} See id.
\textsuperscript{98} See id. at 137 n.36. Professor Frank I. Michelman, in an article that predates \textit{Penn Central} but upon which \textit{Penn Central} relied in its reasoning, justifies this inequity to the extent that such a measure protects the special kind of harm that comes from the psychological impact a property owner suffers when her actual, legally distinct expectations are thwarted seemingly as the result of unprincipled exploitation. See Michelman, supra note 68, at 1229–31.
\textsuperscript{99} See Penn Cent., 438 U.S. at 129, 136, 138 n.36.
\textsuperscript{100} See id. at 115, 138.
panded the denominator beyond the bounds of the property in question to encompass additional property held by the same owner.\footnote{See id. at 138; see also Lucas, 505 U.S. at 1016 n.7 (describing the Penn Central approach as “extreme” and “unsupportable”); Rose, supra note 77, at 567–68 (describing Penn Central’s “expansive” measure of diminution of value against all of the Company’s property).}

B. Rejection of Conceptual Severance and Temporal Severance, and the Expansion of Regulatory Takings Law Beyond Real Property

Penn Central, in which the Supreme Court emphasized that regulatory takings analyses would look to the parcel as a whole as the appropriate denominator, effectively rejected physical severance.\footnote{See Penn Cent., 438 U.S. at 130. In addition to the holding in Penn Central, the event that prevented physical severance from creating the same confusion that occurred in the area of conceptual severance was the holding in Loretto v. Teleprompter Manhattan CATV Corp., which declared all permanent physical invasions to be per se takings. See Loretto v. Teleprompter Manhattan CATV Corp., 458 U.S. 419, 434–35 (1982). It does not matter with respect to which parcel the size of the taking is measured if every physical invasion is a taking, no matter how small. See id.} Conceptual severance, however, remained a troubling problem despite the Court’s attempt to emphasize the importance of measuring the “parcel as a whole.”\footnote{See Penn Cent., 438 U.S. at 130–31.} The difficulty in determining the appropriate denominator only grew as claimants began to argue for temporal severance as a specific instance of conceptual severance.\footnote{See Tahoe-Sierra, 535 U.S. at 320; First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 318–19 (1987); San Diego Gas & Elec. Co. v. City of San Diego, 450 U.S. 621, 657 (1980) (Brennan, J., dissenting).} In 2002, in Tahoe-Sierra Preservation Council, Inc. v. Tahoe Regional Planning Agency, landowners argued that they should be compensated under the Lucas “total economic loss” per se rule because they temporarily lost all economic value in their land during a thirty-two month development moratorium.\footnote{Tahoe-Sierra, 535 U.S. at 320.} The Court dismissed their argument, attempting to declare an end to all severance—physical, conceptual, and temporal—and thus rein in the denominator problem that had grown so troublesome.\footnote{See id. at 326–27; see also Thomas E. Roberts, Regulatory Takings in the Wake of Tahoe-Sierra and the IOLTA Decision, 35 Urb. Law. 759, 759–63 (2003) (decribing Tahoe-Sierra’s resolution of the question of severance).} Emphasizing Penn Central’s end to physical severance, the Court went on to deny conceptual severance, declaring that the destruction of one strand from the bundle of property rights is not a
taking. In addition, *Tahoe-Sierra* explicitly rejected temporal severance. The Court declined to apply the *Lucas* per se rule to the temporary loss of all economic use of the property, turning instead to the multifactor test laid out in *Penn Central*.

In 1982, in *Loretto v. Teleprompter Manhattan CATV Corp.*, the Supreme Court created yet another per se rule as an exception to the *Penn Central* multifactor test. In *Loretto*, a New York City law required that apartment building owners permit cable companies to lay cable along the outside of their buildings. The space occupied by the cable and junction box on Jean Loretto’s building was estimated to be approximately one square foot. The Court relied heavily on the traditional strength of property owners’ right to exclude, characterizing it as one of the most important rights in the “bundle of sticks.” Because of the significance of the right to exclude, the Court declared that any permanent, physical invasion of property, no matter how small, is a taking. Even the minimal imposition of a wire laid across a building triggered the protections of the Takings Clause, because it resulted in a permanent physical imposition which deserved the strongest protection of the Takings Clause.

In the years since *Pennsylvania Coal*, regulatory takings cases have expanded beyond protecting interests in land. Recent cases have considered requiring compensation for required disclosure of trade secrets and retroactive liability for employee health benefits, among other things. Some of the most challenging applications of the Takings Clause to personal property involve government appropriations.

---

107 See *Tahoe-Sierra*, 535 U.S. at 327.
108 See id. at 331–32.
109 Id. at 332.
110 See *Penn Central*, 458 U.S. at 432.
111 Id. at 423.
112 Id. at 438 n.16.
113 See id. at 426.
114 See id. at 434.
115 See *Loretto*, 458 U.S. at 438.
117 See *Ruckelshaus*, 467 U.S. at 1003–04.
118 See *E. Enters.*, 524 U.S. at 537.
of money.\textsuperscript{119} The Supreme Court has applied the Takings Clause to government appropriations of money in at least three cases,\textsuperscript{120} but some commentators question whether this is a reasonable or workable use of the Takings Clause.\textsuperscript{121} There are several arguments against using the Takings Clause to prohibit government appropriations of money, but perhaps the most constitutionally compelling one is that if courts applied the Takings Clause to appropriations of money, then the plenary power of the government to tax would be confused with its limited ability to appropriate property without compensation.\textsuperscript{122} An investigation into the circumstances under which the Takings Clause provides the appropriate framework to analyze appropriations of money is beyond the scope of this Note.\textsuperscript{123} Relying on those cases which have already applied the Takings Clause to monetary exactions, this Note assumes the general applicability of the Takings Clause to government appropriations of money.\textsuperscript{124}

\begin{footnotesize}

\textsuperscript{120} See Brown, 538 U.S. at 231–35; Phillips, 524 U.S. at 163–68; Webb’s, 449 U.S. at 160–64.

\textsuperscript{121} See Epstein, supra note 68, at 99–100, 283–305 (arguing that the Takings Clause applies equally to taxes and appropriations and regulations of real property); Robert Brauneis, Eastern Enterprises, Phillips, Money, and the Limited Role of the Compensation Clause in Protecting Property “in Its Larger and Juster Meaning,” 51 Ala. L. Rev 937, 938 (2000) (rejecting the bundle of rights metaphor, and arguing that the Takings Clause should protect ordinary objects and not legal rights); Michael Allen Wolf, Taking Regulatory Takings Personally: The Perils of (Mis)Reasoning by Analogy, 51 Ala. L. Rev. 1355, 1361–62 (2000) (arguing that the Takings Clause should not apply to government exactions of money, and that instead the constitutionality of such exactions should be decided under the Due Process or Equal Protection Clauses).

\textsuperscript{122} See Brauneis, supra note 121, at 944–47. Brauneis argues that only specific liabilities imposed by the government, associated with identifiable pieces of property, should be subject to the Takings Clause, and that general monetary liabilities, like those imposed in Eastern Enterprises, should be analyzed under the ability of the government to tax. See id.

\textsuperscript{123} For a more in depth discussion of this problem, see Epstein, supra note 68, at 283–305; Brauneis, supra note 121, at 947–48; Wolf, supra note 121, at 1356–63.

\textsuperscript{124} See Brown, 538 U.S. at 231–35; E. Enters., 524 U.S. at 529–37; Phillips, 524 U.S. at 163–68; Webb’s, 449 U.S. at 160–64. The case for the application of the Takings Clause in Givens v. Alabama Department of Corrections was strengthened by the fact that in that case, the interest was taken in addition to a statutorily limited percentage appropriated as a fee. See Webb’s, 449 U.S. at 162; Givens v. Ala. Dep’t of Corr., 381 F.3d 1064, 1065 (11th Cir. 2004). This prevented the Department of Corrections from recharacterizing the appropriation of the interest as a fee or tax. See Webb’s, 449 U.S. at 158; Givens, 381 F.3d at 1065.
\end{footnotesize}
C. Applying Regulatory Takings Law to Takings of Interest Instead of the Severed “Interest Follows Principal” Approach

Courts seem to pay little attention to the history of regulatory takings in their analyses of the inmate interest cases.\textsuperscript{125} Each court presented with the question has relied entirely on the “interest follows principal” rule to guide the discussion, weighing the relative strength of that rule against the common law rule that prisoners hold minimal property rights in the fruits of their labor.\textsuperscript{126} The “interest follows principal” rule, however, artificially separates the ownership of principal and the earned interest in a way that is surprisingly similar to the now discarded notion of conceptual severance.\textsuperscript{127} The right to earn interest on principal, and to direct that interest to designated beneficiaries, is just one right from the bundle of rights that represents ownership of the principal.\textsuperscript{128} Put in a slightly different way, courts could approach the issue as one of a temporary taking of the principal rather than a total taking of the interest, as Justice Breyer suggested in his dissenting opinion in \textit{Phillips}.\textsuperscript{129} If courts employed that analysis, the practice of focusing on whether an inmate “owns” the interest earned distinct from the principal would be revealed as equivalent to temporal severance, with the interest retained by the department of corrections representing the value of the temporary ownership of the principal.\textsuperscript{130} Courts would use the interest lost as only one measure of the economic value of owning the principal for a period of time.\textsuperscript{131} If the courts looked at the ownership of the “parcel as a whole” both temporally and conceptually, instead of splitting their discussion into distinct inquiries into principal and interest, a very different analysis would result.\textsuperscript{132}

\textsuperscript{125} See \textit{Givens}, 381 F.3d at 1068; \textit{Washlefske}, 234 F.3d at 185–86; \textit{Schneider}, 151 F.3d at 1199.

\textsuperscript{126} See \textit{Givens}, 381 F.3d at 1068; \textit{Washlefske}, 234 F.3d at 185–86; \textit{Schneider}, 151 F.3d at 1199.

\textsuperscript{127} See \textit{Givens}, 381 F.3d at 1068; \textit{Washlefske}, 234 F.3d at 185–86; \textit{Schneider}, 151 F.3d at 1199.

\textsuperscript{128} See Webb’s, 449 U.S. at 164.

\textsuperscript{129} See \textit{Phillips}, 524 U.S. at 180 (Breyer, J., dissenting). Oddly, although Justice Breyer suggests the temporary taking of the principal as the starting point for his analysis, he seems to assume that even so, the appropriate rule in most cases is still that “interest follows principal.” See \textit{id}.

\textsuperscript{130} See \textit{id}; cf. \textit{United States v. General Motors Corp.}, 323 U.S. 373, 382–83 (1945) (finding the compensation due after a temporary taking of a warehouse includes the rental value of the property for the period of time during which it was taken).

\textsuperscript{131} See \textit{Tahoe-Sierra}, 535 U.S. at 328–30; \textit{General Motors}, 323 U.S. at 381–83.

\textsuperscript{132} See \textit{infra} notes 133–238 and accompanying text.
III. Analysis: Applying Regulatory Takings Law to Find No Taking of Interest

Abandoning the artificial, conceptual severance of interest and principal results in the prisoner interest cases appearing as instances of temporary regulatory limits on principal, which prevent the owners from using principal to earn interest.133 This analysis replaces the current analysis which sees a potential taking of two distinct properties, the conceptually severed principal and its interest, and its reliance on the common law “interest follows principal” mantra.134 One of the first hurdles to this analysis is determining which of the several established judicial tests of regulatory takings provides the most appropriate and consistent framework.135 Complicating the question of which test to use in the case of the prisoner interest cases is the unusual circumstance of applying the Takings Clause to money and not real property.136 This is because, in such instances, the courts’ reliance on conceptual severance and the “interest follows principal” rule resulted either in finding no property interest to trigger a takings analysis or the total interest to be completely taken, so that the court never identified the appropriate test.137

A. Determining Which Regulatory Takings Test Is the Most Appropriate to Apply to Takings of Interest

The Penn Central Transportation Co. v. City of New York multifactor test, the Lucas per se rule, and the Loretto per se rule are the most relevant tests within the regulatory takings jurisprudence.138 In 2003, in Brown v. Legal Foundation of Washington, the Supreme Court decided

133 See Phillips v. Wash. Legal Found., 524 U.S. 156, 180 (1998) (Breyer, J., dissenting); cf. United States v. General Motors Corp., 323 U.S. 373, 382–83 (1945) (finding the compensation due after a temporary taking of a warehouse includes the rental value of the property for the period of time during which it was taken).


135 See Brown, 538 U.S. at 233–35.

136 See Brauneis, supra note 121, at 937–38; Wolf, supra note 121, at 1355–57.

137 See Givens, 381 F.3d at 1070; Washlefske, 234 F.3d at 185–86; Schneider, 151 F.3d at 1201.

the second of the two IOLTA cases by analogy to “the kind of per se taking that occurred in Loretto.” The Court did so because it was deciding whether the loss of interest was a taking, without reference to the use of the principal. In Eastern Enterprises v. Apfel, in contrast, a plurality of the Supreme Court found that a general liability to pay the government is not the kind of permanent physical occupation that invokes the Loretto per se rule. The difference in the Court’s decision to apply the Loretto per se rule in the two cases might be because in Brown, the Court dealt with the appropriation of specific funds and in Eastern Enterprises, the Court dealt with the imposition of a general monetary obligation. Although the distinction between general and specific obligations may be useful to distinguish between those monetary exactions which should be treated as a tax and those which should be analyzed under the Due Process Clause, that distinction is less useful in determining whether an appropriation of money is like a permanent, physical invasion. The right to exclude formed the basis, in large part, for the Loretto per se rule. In the case of monetary takings, however, there is no question of a right to exclude; only rights to disposition and use are implicated. Taking into account the vague language surrounding the mention of Loretto in Brown, it does not make sense to use Loretto’s physical invasion test, which was developed to protect the right to exclude others from land.

The “exclusive possession” justification for the Loretto rule highlights one of the interesting wrinkles of applying the Takings Clause to money, and one of the reasons why money may be less protected by the Takings Clause than land or personal property. The most protected element from the bundle of rights with respect to property in

---

139 See Brown, 538 U.S. at 235. The Court went on to find that no compensation was due despite the per se taking because the interest lost had no economic value to the owner of the principal. See id. at 239–40.
140 See id. at 234.
142 See Brown, 538 U.S. at 235; E. Enters., 524 U.S. at 530; see also Brauneis, supra note 121, at 937 (discussing the distinction between general liabilities, like taxes, and specific property interests, which are governed by the Takings Clause).
143 See Brauneis, supra note 121, at 943–44 (arguing that specific and general obligations deserve different treatment under the Constitution).
144 See Loretto, 458 U.S. at 435; see also Radin, supra note 86, at 1678–80.
145 See Loretto, 458 U.S. at 435; see also Radin, supra note 86, at 1678–80.
146 See Brown, 538 U.S. at 235.
147 See supra notes 144–46 and accompanying text.
regulatory takings cases is the right to exclude. In questions of taking money, however, the right to exclude has no meaning. The only meaningful property rights to protect with respect to money are the right to use and dispose of it. Although certainly worthy of protection, the Supreme Court has not protected these rights as strongly as the right to exclude.

B. The Lucas Per Se Test Is Not Appropriate for Takings of Money

The Lucas per se test presents an attractive alternative to the Loretto test when applied to the appropriations of money. The Lucas test finds a per se taking when a property owner loses all economically beneficial or productive use of land. Rather than protecting the exclusive possession of land, as in Loretto, the Lucas test protects the use of land, which is a more meaningful right with respect to money. On the one hand, if the interest and principal in the prisoner cases were considered separately, then because the departments of corrections completely appropriated the interest, the “no economically viable use” interpretation of the loss of interest becomes appealing. On the other hand, when the test is applied to the money “as a whole,” that is, to the interest and principal together, application of this test appears less appropriate because some economically beneficial use remains for the principal.

Application of the Lucas test requires overcoming two hurdles. First, the Supreme Court limited the Lucas rule to real property, explic-
itly rendering it inapplicable to personal property. While at first glance that appears to provide a strong argument for declining to apply the Lucas rule to appropriations of money, the justification for the Court’s limitation suggests a potentially reasonable extension to appropriations or limitations on use of money. The Court relied on the understanding that owners of personal property hold their property with the expectation that government regulations may someday render it valueless. According to the Court, however, owners of real property have no such understanding, and expect that regulations may diminish, but not erase, the value of their land. The viability of an extension of the Lucas rule from land to money then turns on whether owners of money hold their money with the expectation that government regulation may cause their money to become worthless, or merely less valuable. The answer to whether extension of the Lucas rule is reasonable is beyond the scope of this Note. The result is the same whether or not the ownership of principal and interest are considered separately, as when the two are conceptually severed, or as part of a unified “parcel as a whole.” Either the Lucas per se rule is inappropriate to apply to either principal or interest, leading to application of the Penn Central multifactor test, or Lucas would be potentially appropriate for both, leading to the second hurdle described below.

Assuming that the Lucas test can provide an appropriate framework for analyzing regulatory takings of money, the test also requires showing that the property owner has lost “all economically beneficial or productive use” of the property. In this inquiry, a difference emerges between the analysis that severs interest and principal and one that joins the two. Any Lucas argument that relies upon the fact that the inmates have lost all economically viable use of their interest also relies on the conceptual severance of the interest from the principal. Under a “parcel as a whole” approach, in contrast, the rele-

158 See Lucas, 505 U.S. at 1027–28. The Court limited the “no economically viable use” per se rule to real property in part because the Court had already held that denying the ability to sell property at all was not necessarily a taking. See id.; Andrus, 444 U.S. at 66–67.
159 See Lucas, 505 U.S. at 1027–28.
160 See id.
161 See id.
162 See id.
163 See id.
164 See Lucas, 505 U.S. at 1027–28.
165 See id. at 1015–16.
166 See id. at 1015.
167 See id. at 1016 n.7.
168 See Brown, 538 U.S. at 234–35; Webb’s, 449 U.S. at 164; Schneider, 151 F.3d at 1201.
vant inquiry is whether the inmates retain any economically viable use of the principal or interest.\textsuperscript{169} The ability to earn interest represents only one use of the principal.\textsuperscript{170} In each of the prisoner interest cases the inmate is at least permitted to retain the principal after leaving prison, so the deprivation imposed by the regulation is at most a temporary total economic deprivation.\textsuperscript{171} Under the Supreme Court’s decision in Tahoe-Sierra, even a total economic deprivation, if it is only temporary, does not automatically trigger a per se taking.\textsuperscript{172} In Washlefske and Schneider, inmates were able to spend principal from prison accounts in the prison commissary and so retained some economically viable use of the principal even while in prison.\textsuperscript{173} In Washlefske, inmates were permitted to send the principal to recipients outside the prison, including for deposit in interest-bearing accounts.\textsuperscript{174} Thus, inmates retained at least some use of their principal.\textsuperscript{175} Because inmates retain significant rights to use and dispose of their principal during and after incarceration, application of the Lucas per se test would fail for lack of a total economic deprivation.\textsuperscript{176} This result would differ if the analysis viewed the loss of interest independent from the use of the principal.\textsuperscript{177}

C. Applying the Penn Central Multifactor Test

Discarding both the Lucas and the Loretto per se rules, the remaining test is the Penn Central multifactor approach.\textsuperscript{178} As articulated by the Supreme Court in Penn Central, the multifactor approach imposes no set formula but rather requires a series of “ad hoc, factual inquiries” into the “particular circumstances of that case.”\textsuperscript{179} Among the relevant factors, the Court lists the character of the governmental action, the economic impact of the regulation, and its interference with distinct, investment-backed expectations.\textsuperscript{180}

\textsuperscript{169} See Lucas, 505 U.S. at 1016 n.7.
\textsuperscript{170} See Phillip’s, 524 U.S. at 180 (Breyer, J., dissenting); Webb’s, 449 U.S. at 164.
\textsuperscript{171} See Givens, 381 F.3d at 1065; Washlefske, 234 F.3d at 181; Schneider, 151 F.3d at 1195.
\textsuperscript{173} See Washlefske, 234 F.3d at 181; Schneider, 151 F.3d at 1195.
\textsuperscript{174} See Washlefske, 234 F.3d at 181.
\textsuperscript{175} See id.
\textsuperscript{176} See Tahoe-Sierra, 535 U.S. at 330; Lucas, 505 U.S. at 1019.
\textsuperscript{177} See Brown, 538 U.S. at 235; Webb’s, 449 U.S. at 164; Schneider, 151 F.3d at 1201.
\textsuperscript{178} See Tahoe-Sierra, 535 U.S. at 334; Penn Cent., 438 U.S. at 124.
\textsuperscript{179} Penn Cent., 438 U.S. at 123–24.
\textsuperscript{180} See id. at 124.
1. The Character of the Government Action Prong

The Court first described the character of the governmental action prong in *Penn Central*, explaining that a regulation that imposed a physical invasion on a property owner was worse than a regulation that “adjusted the benefits and burdens of economic life to promote the common good.” 181 Four years later, *Loretto*, which found a per se taking in all cases of a permanent physical invasion, supplanted the physical invasion element of this factor. 182 The other element of the “character” prong as first described in *Penn Central* is whether or not the regulation in question merely “adjusts benefits and burdens” for the common good. 183 For the most part, the public use requirement of the Takings Clause does not determine the outcome of the “character” factor. 184 In more recent cases, courts interpret the “common good” element of the character of the governmental action to include whether the regulation in question was imposed retroactively, whether the regulation targeted a small number of people or entities to bear a communal burden, as well as the importance of the government purpose the taking served. 185

In the prisoner interest cases, the character of the governmental action prong of the multifactor test results in either a neutral outcome or a finding of no taking. 186 The laws foreclosing the ability to earn interest do not represent a physical invasion, and so fail under that element of the character of the government action prong in the initial *Penn Central* articulation of the test. 187 The laws also pass the low hurdle of appropriately “adjusting the benefits and burdens of economic life,” a hurdle which differs little from the “public use” requirement of any government taking. 188

---

181 See id.; see also Fischel, *supra* note 67, at 50 (noting that the court provided little guidance in interpreting the broad parameters laid out by this prong).
182 See *Loretto*, 458 U.S. at 434–35; see also Peterson, *supra* note 76, at 1317–18 (explaining that before *Loretto*, courts used the ‘character’ prong to ask whether there was a physical invasion).
183 See *Penn Cent.*, 438 U.S. at 124.
184 See Peterson, *supra* note 76, at 1318–19.
185 See *E. Enters.*, 524 U.S. at 537; Bass Enters. Prod. Co. v. United States, 381 F.3d 1360, 1370 (Fed. Cir. 2004); Appolo Fuels, Inc. v. United States, 381 F.3d 1338, 1350–51 (Fed. Cir. 2004); see also Peterson, *supra* note 76, at 1317–19 (arguing that recent developments in the character of the governmental action prong are moving towards incorporating the retroactive or targeted nature of the action).
186 See infra notes 187–92 and accompanying text.
187 See *Penn Cent.*, 438 U.S. at 124. But see *Brown*, 538 U.S. at 235 (suggesting the *Loretto* per se rule for takings of interest).
188 See *Penn Cent.*, 438 U.S. at 124; see also Peterson, *supra* note 76, at 1318–19.
this factor could strengthen the Departments of Corrections’ argument because the regulation required the interest to be deposited in a fund that went toward the benefit of the inmates.\textsuperscript{189} Even under the recent expansion of this element, the regulations in question are not retroactive and do not appear to unfairly target a small group.\textsuperscript{190} The character of the government action prong of the multifactor test, therefore, appears to weigh against the inmates.\textsuperscript{191} This result remains unchanged whether or not the interest is conceptually severed from the principal or the two are treated as a single parcel, because the relative magnitude of the regulation’s impact is irrelevant to the outcome of this factor.\textsuperscript{192}

2. The Economic Impact Prong

The economic impact prong and the interference with distinct, investment-backed expectations prong of the multifactor test have evolved through several interpretations since the Supreme Court described them in 1978.\textsuperscript{193} In the Court’s initial introduction of the multifactor test in \textit{Penn Central}, the economic impact prong included the investment-backed expectation inquiry as one element.\textsuperscript{194} In later cases, however, the Court separated the two into two distinct factors of a three-factor test.\textsuperscript{195} The economic impact prong, when applied independently from the investment-backed expectations prong, in accordance with courts’ most recent practice, measures only the dollar amount of the economic loss suffered by the property owner.\textsuperscript{196}

\textsuperscript{189} See Washlefske, 234 F.3d at 181; Schneider, 151 F.3d at 1195.
\textsuperscript{190} See Givens, 381 F.3d at 1065–66; Washlefske, 234 F.3d at 181–82; Schneider, 151 F.3d at 1195–96.
\textsuperscript{191} See supra notes 181–90 and accompanying text.
\textsuperscript{192} See Webb’s, 449 U.S. at 164–65 (finding a taking as a result of the overwhelming economic impact of the regulation on the loss of interest, but omitting any discussion of the character of the government action).
\textsuperscript{194} See \textit{Penn Cent.}, 438 U.S. at 124.
\textsuperscript{195} See, e.g., \textit{Tahoe-Sierra}, 535 U.S. at 315 n.10; Palazzolo v. Rhode Island, 533 U.S. 606, 617 (2001); \textit{E. Enters.}, 524 U.S. at 529.
\textsuperscript{196} See, e.g., \textit{Brown}, 538 U.S. at 235; \textit{E. Enters.}, 524 U.S. at 529; see also Peterson, supra note 76, at 1325.
In the inmate interest cases, the outcome of the economic impact prong is identical regardless of whether a court separates the interest from the principal or examines the interest and the principal together as an entire parcel. In both analyses, the economic impact of the regulation is the dollar amount of the economic loss of the inmate, which is equivalent to the lost interest. The regulation “took” exactly one right from the prisoner with respect to his property, the right to earn interest on principal earned from a work release program. The dollar amount of the economic impact is difficult to establish because neither the amount of money the inmates would have earned as interest, nor the total amount of interest earned by the departments of corrections, appear in Givens or Schneider. In Washlefske, the court notes that interest from the pooled amounts of all inmate accounts (none of which provided interest to inmates) earned the Department of Corrections $5,479.45 in 1998. In addition to that amount, the prison also earned an average of $59.86 per month from a smaller checking account made up of inmate accounts.

Interpretation of that dollar amount of economic loss—the interest lost, in other words—in the multifactor test is less straightforward. Washlefske contained the amount of money the Department of Corrections earned, but not the corresponding loss to the individual inmate. The economic impact of a potential taking is measured from the perspective of what the claimant lost, not what the government gained. Whether that economic impact is so severe as to rise to the level of a taking depends on how large an impact the regulation imposed in comparison to the remaining rights and economic value the property owner retained in the property. This is the famil-

197 See Brown, 538 U.S. at 234–35; Webb’s, 449 U.S. at 160; Schneider, 151 F.3d at 1119.
198 See Brown, 538 U.S. at 234–35; Webb’s, 449 U.S. at 160; Schneider, 151 F.3d at 1119.
199 See Brown, 538 U.S. at 234–35; Webb’s, 449 U.S. at 164; Schneider, 151 F.3d at 1119.
200 See Givens, 381 F.3d 1064; Schneider, 151 F.3d 1194.
201 Washlefske, 234 F.3d at 181.
202 See id. Washlefske also includes the most detailed information about the amount of money involved from the perspective of an individual inmate. See id. The inmate in this case earned an average of $108.76 per month from being paid $0.90 per hour for labor while in prison. See id. The inmate kept an average monthly closing balance of $87.05 in the two non-interest-bearing accounts maintained in his name by the Department of Corrections. See id.
203 See Fischel, supra note 67, at 50–51; Peterson, supra note 76, at 1325.
204 See Washlefske, 234 F.3d at 181–82.
205 See Brown, 538 U.S. at 236–37.
206 See E. Enters., 524 U.S. at 530–32; Penn Cent., 438 U.S. at 138 n.36 (finding significant the remaining use of Grand Central as a train station in limiting the strength of the economic impact).
iar denominator problem, and it provided the original motivation for the investment-backed expectations prong of the multifactor test.\textsuperscript{207} Courts that have divorced the two prongs have nevertheless been forced to find other methods to weigh the outcome of this prong.\textsuperscript{208}

In all of the prisoner interest cases, inmates were able to use and dispose of the principal in at least some way while in prison.\textsuperscript{209} In some cases, inmates could transfer the principal to recipients outside the prison.\textsuperscript{210} In other cases, inmates could spend the principal within the prison in the commissary.\textsuperscript{211} In all cases, the principal would be returned to the inmate at the conclusion of the prison term.\textsuperscript{212} Although none of the facts relating to the use and disposition of the principal is relevant when considering the deprivation of the interest alone, these facts do become relevant when looking at the entirety of both the principal and the interest as a unified parcel.\textsuperscript{213} Under a unified analysis of the parcel as a whole, and thus viewing earned interest as only one of many uses attaching to ownership of the principal, the loss of that interest is no longer a per se taking because the size of the deprivation of interest appears less significant with respect to the remaining uses of the principal.\textsuperscript{214}

\textsuperscript{207} See Penn Cent., 438 U.S. at 124; see also Michelman, supra note 68, at 1229–34 (using the distinctly perceived interest of the property owner (or, the owner’s “investment-backed expectations”) to define the boundaries of the affected parcel for purposes of measuring the extent of harm).

\textsuperscript{208} See E. Enters., 524 U.S. at 530–32 (finding significant the lack of proportionality between the economic impact and the claimant’s participation in the affected regulatory scheme).

\textsuperscript{209} See Givens, 381 F.3d at 1065 (concerning prisoners able to make withdrawals from the “PMOD” account); Washlefske, 234 F.3d at 181 (explaining that money from “spend accounts” in prison may be spent in the prison commissary, and inmates may send money outside the prison to be invested in interest-bearing accounts, but at least $25 must remain in a “hold account”); Schneider, 151 F.3d at 1195 (explaining that only money from a non-interest-bearing “ITA” can be spent in the prison commissary, and an interest-bearing “IPSA” may only be established if there is at least $25 in an “ITA”).

\textsuperscript{210} See Washlefske, 234 F.3d at 181; Schneider, 151 F.3d at 1195.

\textsuperscript{211} See Washlefske, 234 F.3d at 181; Schneider, 151 F.3d at 1195.

\textsuperscript{212} See Givens, 381 F.3d at 1070; Washlefske, 234 F.3d at 181; Schneider, 151 F.3d at 1195.

\textsuperscript{213} See Brown, 538 U.S. at 234–35; Webb’s, 449 U.S. at 164; Schneider, 151 F.3d at 1201 (all three cases looking at the loss of interest without considering the remaining use of principal).

\textsuperscript{214} See Andrus, 444 U.S. at 65–66 (finding that loss of one “stick in the bundle,” the right to disposition, does not necessarily constitute a taking).
3. The Interference with Distinct Investment-Backed Expectations Prong

The final prong of the *Penn Central* multifactor test is the extent to which the regulation interferes with distinct, investment-backed expectations.\(^{215}\) This prong of the test has caused courts considerable difficulty over the years, so much so that even the term itself has evolved.\(^{216}\) Professor Frank I. Michelman first introduced the term in his article, describing it as a “distinctly perceived, sharply crystallized, investment-backed expectation,”\(^{217}\) and Justice Brennan adopted it in *Penn Central* as “distinct investment-backed expectations.”\(^{218}\) In later Supreme Court cases, the inquiry became one of “reasonable investment-backed expectations.”\(^{219}\) The test initially measured whether the economic impact of a regulation had gone too far.\(^{220}\) In recent years, though, the prong has developed a life of its own, sometimes taking on new meaning, and sometimes being ignored entirely.\(^{221}\) A modern interpretation of distinct or reasonable investment-backed expectations involves the amount of notice afforded the property owner (less

---

\(^{215}\) See *Penn Cent.*, 438 U.S. at 124.

\(^{216}\) See *Fischel*, *supra* note 67, at 50–51; *Peterson*, *supra* note 76, at 1320–25; see also *Givens*, 381 F.3d at 1070; *Washlefske*, 234 F.3d at 181; *Schneider*, 151 F.3d at 1195.

\(^{217}\) Michelman, *supra* note 68, at 1233.

\(^{218}\) *Penn Cent.*, 438 U.S. at 124.


\(^{220}\) See *Penn Cent.*, 438 U.S. at 124; Michelman, *supra* note 68, at 1233. Professor Michelman’s method for determining when compensation should be due required determining “demoralization costs” which he defined as

the total of (1) the dollar value necessary to offset disutilities which accrue to losers and their sympathizers specifically from the realization that no compensation is offered, and (2) the present capitalized dollar value of the lost future production . . . caused by demoralization of uncompensated losers, their sympathizers, and other observers disturbed by the thought that they themselves may be subject to similar treatment on some occasion.

Michelman, *supra* note 68, at 1214. Interference with “distinctly perceived, sharply crystallized, investment-backed expectation” leads to high demoralization costs. See *id.* at 1229–34. For a refreshingly clear explanation of this analysis, see *Fischel*, *supra* note 67, at 142–51.

protection is provided for those in highly regulated industries, and more protection is provided for retroactive regulations) and any reasonable reliance the owner had on the absence of the regulation.222

In this more recent incarnation, divorced from the economic impact of the regulation, the investment-backed expectations prong of the multifactor test fails to fulfill its initial purpose—how to measure the relative importance of economic loss on a property owner by determining a comparative property loss.223 For example, a million-dollar economic impact may, in isolation, appear significant, but not if it occurs with respect to a piece of property valued at several billion dollars.224 Similarly, an economic impact of a few hundred dollars may seem trivial to a court unless it is revealed that it is lost from a piece of property worth only one thousand dollars.225 This, of course, is a version of the “denominator problem” mentioned above.226

Applying the earlier version of the investment-backed expectations test to the prisoner interest cases, the question becomes to what extent does a prisoner perceive the right to earn interest as a distinct right, crystallized and worthy of protection.227 One indicator of such a right would be an owner’s independent, “distinct” investment in such a right, but other indicators are possible.228 For example, the legal treatment of principal/interest cases, which analyzes the two parts separately, would argue for a common understanding of the right to earn interest as a distinct, protectable interest.229 The right to earn interest on principal, however, is not ordinarily considered a separate, protected right attached to property ownership.230 The air rights at issue in Penn Central, for example, are more easily thought of as dis-

222 See E. Enters., 524 U.S. at 532–36 (discussing retroactivity under reasonable investment-backed expectations prong); Ruckelshaus v. Monsanto Co., 467 U.S. 986, 1007 (1984) (finding no reasonable investment-backed expectation because plaintiff had sufficient notice of the regulation); Appolo Fuels, 381 F.3d at 1349–50 (finding notice and a highly regulated industry relevant to the reasonable investment-backed expectation analysis).

223 See Penn Cent., 438 U.S. at 124; Michelman, supra note 68, at 1229–34.

224 See Penn Cent., 438 U.S. at 138 n.36 (finding no taking because the property could still be profitably used as a railroad station); id. at 141 (Rehnquist, J., dissenting) (describing the lost revenue resulting from the regulation as at least $3 million per year); see also Rose, supra note 77, at 566–69.

225 See Michelman, supra note 68, at 1190; Rose, supra note 77, at 566–69.

226 See supra notes 83–101 and accompanying text.

227 See Penn Cent., 438 U.S. at 124–25; Michelman, supra note 68, at 1233.

228 See Penn Cent., 438 U.S. at 124–25; Michelman, supra note 68, at 1233.

229 See Phillips, 524 U.S. at 163–66; Webb’s, 449 U.S. at 160–62; Givens, 381 F.3d at 1068; Washlefske, 234 F.3d at 185–86; Schneider, 1199–201.

230 See Phillips, 524 U.S. at 170 (discussing interest as government created value).
tinct from the ownership of the earth underneath, both legally (with the invention of “transferable development rights”) and physically; yet the Supreme Court declined to find that the loss of value of those air rights involved separable, distinct, investment-backed expectations.\footnote{See Penn Cent., 438 U.S. at 130.} Under a unified, “parcel as the whole” analysis, the investment-backed expectations prong of the \textit{Penn Central} test would probably weigh in favor of the department of corrections, reasoning that the economic impact of the lost interest is a relatively insignificant imposition on inmates given the remaining uses of the principal.\footnote{See \textit{id.} at 137–38.}

An analysis of the investment-backed expectations prong using more recent court interpretations based on notice or reliance is also likely to weigh in favor of no taking.\footnote{See, e.g., \textit{E. Enters.}, 524 U.S. at 533–36; \textit{Ruckelshaus}, 467 U.S. at 1005–08.} Inmates are in a highly regulated environment, calling into question any reasonable expectation they may have for earning interest.\footnote{See \textit{Ruckelshaus}, 467 U.S. at 1007; \textit{Appolo Fuels}, 381 F.3d at 1349.} In one sense, inmates’ reliance on the interest might be great because they have no other options.\footnote{See \textit{Schneider}, 151 F.3d at 1196.} A prison is such a heavily regulated environment, however, that a court is likely to find that reliance unreasonable.\footnote{See \textit{Ruckelshaus}, 467 U.S. at 1007; \textit{Appolo Fuels}, 381 F.3d at 1349; see also \textit{Eagle}, supra note 193, at 443–44.} Furthermore, notice problems are minimal because the regulations are not retroactive and inmates are informed of the regulation before depositing money in the prison accounts.\footnote{See \textit{E. Enters.}, 524 U.S. at 532–36; \textit{Appolo Fuels}, 381 F.3d at 1350.} Thus, under either early or late versions of the investment-backed expectations prong, a court would likely find that the interference with prisoners’ expectations is relatively minor when considered in conjunction with the remaining benefits of ownership of principal, leading this prong to weigh against finding a taking.\footnote{See supra notes 215–37 and accompanying text.}

\textbf{Conclusion}

Courts that rely on the “interest follows principal” rule incorporate the idea of conceptual severance into regulatory takings law. This conflicts with the development of regulatory takings decisions, which have consistently attempted to disavow the practice of conceptual severance. The most explicit disavowal of conceptual severance has come in decisions concerning attempted temporal severance, yet the uni-
versally applied “interest follows principal” rule is an example of temporal severance. Although such a rule may have been useful in the eighteenth century, and is still useful in many areas of the law, in the area of regulatory takings, it is no longer needed. Not only would the disavowal of the “interest follows principal” rule be consistent with the rest of regulatory takings law, but this is one circumstance in which an application of regulatory takings doctrine would provide a clearer, rather than more ambiguous, outcome.

Rebecca Rogers
PUBLIC EMPLOYEE FREE SPEECH: THE POLICY REASONS FOR REJECTING A PER SE RULE PRECLUDING SPEECH RIGHTS

Abstract: Public employees do not enjoy the same free speech rights under the First Amendment as do ordinary citizens. The Fourth Circuit Court of Appeals recently adopted a virtual per se rule precluding free speech rights for public employees while they are performing ordinary job duties. This Note argues that such a per se rule both lacks policy justification and, more importantly, would undermine the purposes of the First Amendment by impeding academic freedom and permitting viewpoint discrimination. Rejecting the per se rule best preserves the free speech rights of public employees and in turn allows them to speak freely about governmental operations—an especially important function of public employees, who are firsthand witnesses to government activities.

Introduction

Although the First Amendment generally protects employees’ free speech rights, the rights of those employed by the government are more limited.\(^1\) Courts, most recently the Ninth Circuit Court of Appeals, have wrestled with how best to define the contours of public employee free speech rights.\(^2\) In 2004, in *Ceballos v. Garcetti*, the Ninth Circuit expressly rejected a per se rule precluding First Amendment protection for public employees performing regular work duties.\(^3\) In doing so, *Ceballos* sustained the view that public employees do not give up all free speech protection as a condition of working for the government.\(^4\) Furthermore, *Ceballos* is significant because it relied upon judicial analyses of free speech in the public employment context, one of the few areas in which courts limit free speech rights under the First Amendment.\(^5\)

---

2. See id.
3. Id. at 1178.
4. See id. at 1174.
5. See id.
The First Amendment to the U.S. Constitution states that “Congress shall make no law . . . abridging the freedom of speech, or of the press, or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.”\(^6\) First Amendment doctrine, however, limits the general principle that the government cannot abridge free speech in several ways.\(^7\) For example, the government may suppress speech if it is directed at inciting or producing imminent lawless action and is likely to incite or produce such action.\(^8\) This formulation is the modern version of the “clear and present danger” doctrine.\(^9\) Additionally, the government may regulate certain types of obscenity and indecency.\(^10\) Libelous speech also is usually impermissible under the First Amendment.\(^11\) Finally, the government possesses limited ability to restrict speech by its employees.\(^12\)

An inherent tension exists between government employers’ need to manage the workplace and the operation of the First Amendment in preventing governmental infringement of free speech.\(^13\) Courts have struggled with the balance of public employees’ free speech rights and government interests for over a hundred years, always recognizing, on the one hand, that full operation of First Amendment free speech in public workplaces would allow impermissible work-

---

\(^6\) U.S. Const. amend. I.

\(^7\) See infra notes 8–12 and accompanying text.

\(^8\) Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (enunciating the test that the government may suppress speech only if it is directed to inciting or producing imminent lawless action and is likely to incite or produce such action).

\(^9\) Schenck v. United States, 249 U.S. 47, 52 (1919) (holding that the defendants violated federal statutes by interfering with military recruitment during wartime and that the speech lacked First Amendment protection because it posed a clear and present danger to national security). Schenck v. United States set out the original clear and present danger test. Id. Brandenburg v. Ohio is the modern precedent on the issue. See 395 U.S. at 447.

\(^10\) See New York v. Ferber, 458 U.S. 747, 764 (1982) (holding that child pornography receives no First Amendment protection so long as the applicable statute banning child pornography explicitly restricts only works that visually portray children in a sexually explicit manner); Miller v. California, 413 U.S. 15, 24 (1973) (formulating a three-factor test to determine if sexually explicit material is obscene).

\(^11\) See N.Y. Times v. Sullivan, 376 U.S. 254, 269 (1964) (holding that in order to comply with First Amendment freedoms, public officials alleging libel must prove actual malice to recover).

\(^12\) See Pickering v. Bd. of Educ., 391 U.S. 563, 568 (1968) (holding that to receive First Amendment protection, a public employee’s interest in speaking must outweigh the public employer’s interest in preventing workplace disruption).

\(^13\) See id.
place disruption. On the other hand, public employees are in a unique position to uncover improprieties within the government due to their firsthand knowledge of governmental operations, and therefore, the First Amendment’s guarantee of a right to petition the government for redress of grievances is an essential right to preserve for public employees.

U.S. courts offered little protection for public employees’ free speech rights before 1968. For example, in 1892, in McAuliffe v. Mayor of New Bedford, the Supreme Judicial Court of Massachusetts rejected the right of public employees to criticize their employers and maintain their employment. Justice Oliver Wendell Holmes is often quoted for the introductory remark to his opinion in that case: “The petitioner may have a constitutional right to talk politics, but he has no constitutional right to be a policeman.” Accordingly, government employees were forced to give up constitutional free speech rights as a condition of public employment.

Courts today provide some First Amendment protection to public employees. Although public employees do not enjoy the same free speech rights at work as ordinary citizens, certain protections are afforded. Whether public employees’ speech is constitutionally protected is determined using several judicial tests. In 2000, in Urofsky v. Gilmore, the Fourth Circuit Court of Appeals decided that if a public employee is performing regular workplace duties and thus acting as a representative of the state, the employee’s speech should not receive First Amendment protection. Such a per se rule would create a new type of analysis for when a public employee speaks while performing job duties. Although the Fourth Circuit justifies its approach by rea-

---

15 See Ceballos, 361 F.3d at 1175.
16 See Washington, 84 F. Supp. at 966; McAuliffe, 29 N.E. at 517; Reagan, 284 S.W.2d at 937; see also Lawrence Rosenthal, Permissible Content Discrimination Under the First Amendment: the Strange Case of the Public Employee, 25 Hastings Const. L.Q. 529, 530 (1998) (describing the history of public employee free speech jurisprudence).
17 See id.; see also Washington, 84 F. Supp. at 966; McAuliffe, 29 N.E. at 517.
18 See infra notes 32–54 and accompanying text.
19 See infra notes 32–54 and accompanying text.
20 See, e.g., Ceballos, 361 F.3d at 1174–76; Urofsky v. Gilmore, 216 F.3d 401, 409 (4th Cir. 2000).
21 See id. at 407.
soning from current caselaw, such an approach potentially could eradicate the use of the current judicial public employee speech tests.\(^\text{25}\)

This Note argues that courts must reject the per se rule precluding First Amendment protection for public employees’ speech made while performing regular employment duties.\(^\text{26}\) Part I of this Note explains the U.S. Supreme Court jurisprudence in the area of public employees’ free speech rights and discusses the relevance of whistleblower statutes.\(^\text{27}\) Part II discusses the concept of a per se rule preventing public employees from claiming First Amendment protection for speech in their regular employment duties.\(^\text{28}\) Part III discusses \textit{Ceballos}, in which the Ninth Circuit expressly rejected a per se rule.\(^\text{29}\) This Part also discusses the status of the per se rule in other circuits and the common policy objections to the per se rule.\(^\text{30}\) Finally, Part IV argues that the purposes of the First Amendment will be undermined unless public employees possess freedom of speech while performing regular employment duties.\(^\text{31}\)

\section*{I. The History of Public Employees’ Free Speech Rights}

\subsection*{A. Two-Tiered Analysis: The Pickering Balancing Test and the Connick Public Concern Test}

The modern approach to public employees’ speech doctrine began in the 1960s when the U.S. Supreme Court addressed the issue of public employees’ free speech rights.\(^\text{32}\) In 1968, in \textit{Pickering v. Board of Education}, the Court articulated a new balancing test to determine whether a public employee’s speech is constitutionally protected.\(^\text{33}\) The plaintiff, a public school teacher, lost his job for writing a letter to a local newspaper alleging irresponsible spending by the school district on sports programs.\(^\text{34}\) The Court determined that the Board of Education had violated the plaintiff’s First Amendment right to free speech.\(^\text{35}\) Us-

\(^{25}\) See id.

\(^{26}\) See infra notes 160–245 and accompanying text.

\(^{27}\) See infra notes 32–69 and accompanying text.

\(^{28}\) See infra notes 70–105 and accompanying text.

\(^{29}\) See infra notes 106–59 and accompanying text.

\(^{30}\) See infra notes 122–59 and accompanying text.

\(^{31}\) See infra notes 160–245 and accompanying text.


\(^{33}\) 391 U.S. at 568.

\(^{34}\) \textit{Id.} at 564.

\(^{35}\) \textit{Id.} at 565.
ing a formula now referred to as the *Pickering* balancing test, the Court held that for a statement by a public employee to receive First Amendment protection, the employee’s interest “as a citizen, in commenting upon matters of public concern,” must outweigh “the interest of the State, as an employer, in promoting the efficiency of the public services it performs through its employees.”

Because teachers are likely to have informed opinions on the operation of schools, the Court reasoned that teachers should be able to speak about how their area of government should allocate funds. Furthermore, the Court concluded that even though some of the plaintiff’s statements were false, those statements nevertheless constituted protected speech because they were not knowingly or recklessly false.

The *Pickering* balancing test is still used in public employee free speech cases, supplemented by an additional preliminary analysis that the Court later added to this test. In 1983, in *Connick v. Myers*, the U.S. Supreme Court sought to clarify its holding in *Pickering*. The *Connick* Court held that as a threshold issue of law, before applying the *Pickering* test, courts must determine whether the public employee speech at issue touched matters of public concern. In *Connick*, the Court held that a public employee did not have a free speech right to prepare a questionnaire for her coworkers where the complaints at issue—including transfer policies, office morale, pressure to vote for certain political parties, and the need for a grievance committee in her government office—were primarily personal grievances, not matters of public concern. The Court reiterated the importance of ensuring that citizens retain fundamental rights while working for the government. The Court indicated that the essential inquiry under *Connick* is to determine whether an analysis of a public employee’s speech can proceed to the *Pickering* test, which depends upon whether the employee spoke “as a citizen upon matters of public concern.”

---

36 Id. at 568.
37 Id. at 572.
38 *Pickering*, 391 U.S. at 572.
40 See *Connick*, 461 U.S. at 146.
41 Id.
42 Id. at 141–42. The Supreme Court stated that the district court and court of appeals had misapplied the *Pickering* balancing test. Id. at 142. The statement at issue, however, only minimally touched upon matters of public concern. Id. at 154.
43 Id. at 147.
rather than “as an employee upon matters only of personal interest.”

Courts now utilize the “public concern” test. The Connick Court provided several guiding statements regarding the operation of the Pickering balancing test. The Connick Court stated that when employee expression does not relate to a matter of political, social, or other concern to the community, the government can manage its offices—and employee speech—as it sees fit without worry of judicial intervention. When a public employee speaks on a matter of public concern, however, the speech is of sufficient importance to receive protection in federal courts.

The Connick Court decided that determining what constitutes a matter of public concern is a question of law, and courts must analyze the “content, form, and context” of the allegedly protected speech in making this determination. Because the Court has restricted protected speech to that addressing matters of public concern, not every statement made in a government office, particularly those regarding workplace grievances that have no societal importance, is subject to constitutional scrutiny. The more an employee’s speech addresses a matter of public concern, however, the greater a showing of workplace disruption is necessary to disallow First Amendment protection under the Pickering balancing test. The Court nevertheless gives great deference to employers and allows preventive measures to disruption when necessary. Thus, as a result of Pickering and Connick, a trial court presented with a public employee’s free speech claim first must consider whether a public employee’s statement touches on a matter of public concern, and then the court must weigh the interests of the government, including preventing workplace disruption, against the em-

44 Id.
45 E.g., Rodgers v. Banks, 344 F.3d 587, 596 (6th Cir. 2003); Baldassare v. New Jersey, 250 F.3d 188, 197 (3d Cir. 2001); Dill v. City of Edmond, 155 F.3d 1193, 1201 (10th Cir. 1998); Fikes v. City of Daphne, 79 F.3d 1079, 1084 (11th Cir. 1996).
46 See Connick, 461 U.S. at 138–54 (citing Pickering for the proposition that a statement by a public employee receives First Amendment protection only if the employee’s interest “as a citizen, in commenting upon matters of public concern,” outweighs “the interest of the State, as an employer, in promoting the efficiency of the public services it performs through its employees”).
47 Id. at 146.
48 Id.
49 Id. at 147.
50 Id. at 149.
51 Connick, 461 U.S. at 152.
52 See id.
ployee’s interest in the speech.\textsuperscript{53} Courts now use the threshold \textit{Connick} public concern test and the \textit{Pickering} balancing test to analyze all public employee free speech cases.\textsuperscript{54}

B. Additional U.S. Supreme Court Commentary on Public Employees’ Free Speech Rights

In the years following \textit{Connick} and \textit{Pickering}, the Supreme Court clarified the operation of the two tests.\textsuperscript{55} For example, in 1987, in \textit{Rankin v. McPherson}, the Court held that public employee speech need not directly relate to one’s employment to receive First Amendment protection.\textsuperscript{56} In that case, the defendant, a constable, terminated the plaintiff, a law enforcement official, when the plaintiff remarked that she hoped assassins would successfully shoot President Ronald Reagan the next time they tried.\textsuperscript{57} The Court held that the plaintiff’s statements were of public concern under \textit{Connick}, and that the employee’s interest in speaking outweighed the employer’s interest in forbidding it under the \textit{Pickering} balancing test; therefore, the speech received First Amendment protection.\textsuperscript{58} The Court reasoned that the plaintiff’s statement did not disrupt the work environment, and that because it did not make her unsuitable to perform her work, no governmental interest existed to balance against the employee’s free speech interest under the \textit{Pickering} test.\textsuperscript{59}

Furthermore, the Supreme Court also has restrained government employers’ right to terminate employees immediately for criticizing the government.\textsuperscript{60} In 1996, in \textit{Waters v. Churchill}, the Court held that a government employer must conduct a reasonable inquiry into an alleged unprotected statement before taking any adverse employment action because of it.\textsuperscript{61} In \textit{Waters}, the Court decided that a public hospital nurse’s interest in criticizing a hospital policy that she thought adversely affected patient care was of public concern under \textit{Connick} and sufficiently outweighed the government employer’s interest in an

\textsuperscript{53} See \textit{id.} at 146; see also \textit{Pickering}, 391 U.S. at 568.
\textsuperscript{54} See, e.g., Rodgers, 344 F.3d at 596; Baldassare, 250 F.3d at 197; Dill, 155 F.3d at 1201; \textit{Fikes}, 79 F.3d at 1084.
\textsuperscript{56} \textit{Id.} at 389.
\textsuperscript{57} \textit{Id.} at 379–80.
\textsuperscript{58} \textit{Id.} at 388–89.
\textsuperscript{59} \textit{Id.} at 388.
\textsuperscript{61} \textit{Id.} at 677.
efficient workplace under *Pickering*. This holding grants some protection for whistleblowers because it means they do not risk immediate termination after reporting wrongdoing, though the Court suggested that additional procedural protections could be given through whistleblower statutes.

C. Additional Category of Free Speech Protection: The Whistleblower Statutes

In addition to the protection accorded to public employees’ free speech rights under *Pickering* and *Connick*, whistleblower statutes also protect federal employees. The major law protecting federal employees’ speech rights is the Whistleblower Protection Act of 1989. The Act provides that federal employees who reveal particular prohibited personnel practices and experience some adverse personnel action as a result may seek redress from a Merit Systems Protection Board. If the public employee can show both that the official taking the personnel action knew of the disclosure and that the personnel action occurred within a period of time such that a reasonable person could conclude that the disclosure was a contributing factor in the personnel action, the employee is entitled to corrective action. Employee revelations protected by the Act include violations of any law, rule, or regulation; gross mismanagement; a gross waste of funds; an abuse of authority; or a substantial and specific danger to public safety or health. Though whistleblower legislation is sometimes criticized as under-inclusive, it does exist as an additional protection for public employees’ free speech rights.

---

62 *Id.* at 675, 677. The *Waters* Court further held that a trial court should decide whether speech is of public concern based upon what content the employer reasonably believed the speech contained, not what content the factfinder determines the speech contained. *Id.* at 677.
63 *Id.* at 674.
65 *See id.*
66 *See § 1221(e)(1).*
67 *See id.*
69 *See infra* notes 228–36 and accompanying text.
II. THE MOVE TOWARD A PER SE RULE AGAINST FREE SPEECH RIGHTS WITHIN A PUBLIC EMPLOYEES’ JOB DUTIES

A. The Distinction Between Private Citizens and Public Employees in the Connick Public Concern Test

The U.S. Supreme Court has not yet decided a case in which a public employee’s speech is part of that employee’s fulfillment of an employment duty.\(^70\) Rather, the Court’s cases have dealt only with extraneous comments made by employees during work, not speech made in memoranda or reports the employee is required to write to fulfill ordinary employment duties while acting on behalf of the government.\(^71\) Whether this type of speech receives First Amendment protection remains uncertain because the federal circuit courts of appeals are divided on the issue.\(^72\)

This uncertainty regarding whether public employee speech made while carrying out one’s employment duties receives constitutional protection may depend upon the interpretation of language in *Connick v. Myers.*\(^73\) *Connick*’s public concern test distinguishes protected speech based on whether a public employee speaks “as a citizen upon matters of public concern” or “as an employee upon matters only of personal interest.”\(^74\) A public employee speaking as a citizen on matters of public concern is constitutionally protected, but an employee speaking on matters of private concern receives no First Amendment protection.\(^75\) Consequently, two questions arise: If the public employee is speaking on a matter of public concern, can the employee still speak as a citizen within the meaning of *Connick* when performing employment duties in the role as employee—for example, while writing reports or speaking publicly as a representative of the government?\(^76\) Moreover, has *Connick* set up a test in which courts must determine separately whether the subject matter touches public concern and whether the employee was speaking as an employee or as

---


\(^71\) See *Connick*, 461 U.S. at 138–54; *Pickering*, 391 U.S. at 563–75.

\(^72\) Compare *Ceballos v. Garcetti*, 361 F.3d 1168, 1178 (9th Cir. 2004), *cert. granted*, 125 S. Ct. 1395 (2005) (holding that no per se rule exists), with *Urofsky v. Gilmore*, 216 F.3d 401, 409 (4th Cir. 2000) (holding that public employees enjoy no free speech rights while speaking in their employee roles).

\(^73\) See *Urofsky*, 216 F.3d at 409.

\(^74\) *Connick*, 461 U.S. at 147.

\(^75\) Id.

\(^76\) See supra notes 70–75; infra notes 78–159 and accompanying text.
a citizen?77 The Fourth Circuit Court of Appeals favors a per se rule precluding First Amendment protection for public employees performing their job tasks (“the per se rule”) and considers the Connick language to create an essential distinction, while the Ninth Circuit Court of Appeals rejects the per se rule and interprets the Connick language more loosely.78 These opposing views are discussed below.79

B. Urofsky v. Gilmore: The Fourth Circuit Moves Toward a Per Se Rule

The Fourth Circuit has adopted a virtual per se rule precluding free speech rights for public employees.80 In 2000, in Urofsky v. Gilmore, the Fourth Circuit held that in determining whether public employee speech receives First Amendment protection, a court first must consider whether the employee spoke as a private citizen or as a public employee; only after this consideration can a court determine whether the speech was of public concern under Connick and then move on to perform the Pickering v. Board of Education balancing test.81

The Urofsky court adopted this virtual per se rule in the context of a case where professors at public colleges and universities brought suit to challenge the constitutionality of a Virginia statute restricting state employees’ access to sexually explicit material on computers owned or leased by the state.82 The court decided the case using the framework of the Pickering and Connick tests and determined that state employees receive no constitutional protection while accessing sexually explicit materials on state-owned computers.83

In applying the Connick public concern test, however, the Urofsky court interpreted the U.S. Supreme Court’s enunciated test—whether the employee spoke “as a citizen upon matters of public concern,” rather than “as an employee upon matters only of personal interest”—as emphasizing the unrelatedness of the unprotected speech at issue to the public employee’s work duties.84 To demonstrate this distinction, the court used the example that if a district attorney must make a public statement regarding an upcoming murder trial, the attorney

77 See supra notes 70–75; infra notes 78–159 and accompanying text.
78 Compare Ceballos, 361 F.3d at 1178 (rejecting a per se rule), with Urofsky, 216 F.3d at 409 (essentially adopting a per se rule).
79 See infra notes 80–159 and accompanying text.
80 See Urofsky, 216 F.3d at 409.
81 Id. at 407.
82 Id. at 404.
83 Id.
84 Id. at 407 (citing the Connick public concern test).
has no constitutional free speech right to disregard a supervisor’s instructions about the speech because the speech is a requirement of the attorney’s employment, even though the content of that speech may be of great public concern. The attorney could, however, write a letter to a newspaper alleging prosecutorial malfeasance—meaning the attorney would be acting as a citizen—and receive First Amendment protection. In drawing this distinction, the court explained that employment duties will often involve speaking on matters of concern to the public and that using the First Amendment to give employees the right to dictate how they perform those work duties—simply because the speech involves a matter of public concern—would be irrational.

_Urofsky_ does not establish a firm per se rule, but if it is applied stringently, public employees speaking in their roles as employees have no free speech rights. The _Urofsky_ court asserted that when speech is not made in a public employee’s capacity as a private citizen and regarding matters of public concern, no First Amendment protection exists. Although the court did not expressly state that a public employee never has a free speech right while performing employment duties, _Urofsky’s_ holding made the role of the speaker dispositive in the public concern analysis under _Connick_. Consequently, _Urofsky_ requires a new two-tiered inquiry when an employee speaks in fulfillment of an employment duty. A court must consider first whether the employee spoke as a citizen or as an employee and then consider whether the subject matter was of public concern.

C. Policy Support for the Per Se Rule

_Urofsky_ highlights multiple policy concerns that underlie the idea of adopting a per se rule, and its viewpoint has also proven influential with other judges. First, public employees speak for the state. Second, providing such speech rights would be inconsistent with whistle-
blower statutes.\textsuperscript{95} Third, the rule would constitutionalize the performance of everyday work duties.\textsuperscript{96}

The first major policy supporting a per se rule is that when public employees act within the scope of their employment, they speak not for themselves but for the state.\textsuperscript{97} As a result, the employee’s role requires regulation because observers could construe statements as official government stances on any given issue, regardless of whether the statements address matters of public concern.\textsuperscript{98} No personal interest in the speech exists, and consequently, no First Amendment right should attach.\textsuperscript{99}

Second, according to the Urofsky viewpoint, legislatures enacted whistleblower statutes solely because such legislation creates a right where no constitutional protection otherwise exists.\textsuperscript{100} In supporting the Urofsky court view in his special concurrence to the 2004 Ninth Circuit Court of Appeals decision in \textit{Ceballos v. Garcetti}, Judge O’Scannlain noted that whistleblower statutes became necessary because the First Amendment did not otherwise protect public employees who revealed governmental misdeeds.\textsuperscript{101} Consequently, legislatures passed laws to protect this speech because it had value but remained constitutionally unrecognized.\textsuperscript{102}

Finally, the Urofsky court reasoned that without a per se rule, a constitutional claim could arise in every action a public employee performs.\textsuperscript{103} Out of their fear of lawsuits, government employers could be forced to tolerate mere workplace grievances or unprofessional language and to avoid making rules and taking disciplinary action.\textsuperscript{104} Because of public employees’ status as representatives of the government, the existence of whistleblower statutes, and the fear of constitutionalizing everyday work duties, some reason that a per se rule against constitutional protection for public employee speech is necessary.\textsuperscript{105}

\textsuperscript{95} See infra notes 100–02 and accompanying text.
\textsuperscript{96} See infra notes 103–04 and accompanying text.
\textsuperscript{97} See \textit{Ceballos}, 361 F.3d at 1189 (O’Scannlain, J., specially concurring).
\textsuperscript{98} See \textit{id.} Judge O’Scannlain argued that performance of a public employment task is state action, and therefore, that speech contained within an employment task bears the seal of the state. See \textit{id.}
\textsuperscript{99} See \textit{id.}
\textsuperscript{100} See \textit{id.} at 1192.
\textsuperscript{101} See \textit{id.}
\textsuperscript{102} See \textit{Ceballos}, 361 F.3d at 1192 (O’Scannlain, J., specially concurring).
\textsuperscript{103} See Urofsky, 216 F.3d at 408.
\textsuperscript{104} See \textit{id.} at 409.
\textsuperscript{105} See supra notes 93–104 and accompanying text.
III. The Rejection of the Per Se Rule

A. Rejecting Urofsky: Ceballos v. Garcetti

Although some courts maintain the necessity of a per se rule precluding free speech rights for public employees performing job duties, others take the position that some constitutional protection for public employees at work necessarily must exist. In 2004, in Ceballos v. Garcetti, the Ninth Circuit Court of Appeals expressly rejected the per se rule. In Ceballos, the plaintiff, a deputy district attorney, was instructed by his superiors to investigate the possibility that a deputy sheriff lied in a search warrant affidavit. The plaintiff sent a memorandum to the head deputy district attorney, who informed the plaintiff that he should revise the memorandum to make it less accusatory of the sheriff. While a hearing on a motion challenging the search warrant was pending, the plaintiff informed opposing counsel that he believed the affidavit contained false statements and that he felt the law compelled him to turn over his memorandum to opposing counsel. The plaintiff’s superiors then instructed him to amend the memorandum to include the statements of only one detective, and as a result, the plaintiff was prevented from telling the court some of his conclusions.

The plaintiff was then demoted, treated in a hostile manner, and denied a promotion for criticizing the deputy sheriff too harshly. He filed a complaint against his employer alleging a violation of his First Amendment rights. The district court dismissed the complaint against the government on Eleventh Amendment grounds and against individual defendants on the grounds of qualified immunity.

On appeal, the Ninth Circuit reversed as to the individual defendants, holding that public employees can have free speech rights while carrying out employment duties. The plaintiff’s employers...
contended that even though exposing a sheriff’s false testimony undoubtedly addresses a matter of public concern, the memorandum that the plaintiff prepared was unprotected speech because he prepared it in fulfillment of an employment responsibility.\textsuperscript{116} Under the employers’ conception of the \textit{Connick v. Myers} public concern test, the plaintiff had acted as an employee and not as a citizen; therefore, they argued, the speech should not receive constitutional protection.\textsuperscript{117}

The court ultimately rejected this argument, reasoning that public employees are in a unique position to contribute to the debate on matters of public concern, and that depriving them of free speech protection when carrying out their employment duties could undermine the maintenance of governmental integrity.\textsuperscript{118} Additionally, the court stated that a per se rule would negatively affect whistleblowers because public employees have a duty to notify their supervisors of wrongdoings.\textsuperscript{119} Moreover, the court noted that if a distinction is made between whether the employee acted as a citizen or as an employee, a public employee could receive protection for reporting some operation of the government to the press but not to supervisors who might be able to remedy the situation.\textsuperscript{120} \textit{Ceballos} thus expressly criticized the Fourth Circuit’s decision in \textit{Urofsky v. Gilmore} and then rejected a per se rule.\textsuperscript{121}

\section*{B. Other Federal Courts of Appeals Expressly or Impliedly Rejecting a Per Se Rule}

Like the Ninth Circuit in \textit{Ceballos}, most circuits have either impliedly or expressly rejected the per se rule.\textsuperscript{122} In 2001, in \textit{Baldassarre v. New Jersey}, the Third Circuit expressly rejected a per se rule.\textsuperscript{123} In that case, a prosecutor’s office demoted and later fired an investigator af-

\textsuperscript{116} \textit{Id.} at 1174.
\textsuperscript{117} \textit{Ceballos}, 361 F.3d at 1174; \textit{see also} \textit{Connick v. Myers}, 461 U.S. 138, 147 (1983).
\textsuperscript{118} \textit{Ceballos}, 361 F.3d at 1175.
\textsuperscript{119} \textit{Id.} at 1176.
\textsuperscript{120} \textit{Id.}
\textsuperscript{121} \textit{Id.} at 1177 n.7; \textit{see} \textit{Urofsky v. Gilmore}, 216 F.3d 401, 409 (4th Cir. 2000).
\textsuperscript{122} \textit{See}, e.g., \textit{Rodgers v. Banks}, 344 F.3d 587, 598–99 (6th Cir. 2003); \textit{Baldassare v. New Jersey}, 250 F.3d 188, 196–97 (3d Cir. 2001); \textit{Lewis v. Cowen}, 165 F.3d 154, 164 (2d Cir. 1999); \textit{Dill v. City of Edmond}, 155 F.3d 1193, 1202–03 (10th Cir. 1998); \textit{Fikes v. City of Daphne}, 79 F.3d 1079, 1084 (11th Cir. 1996).
\textsuperscript{123} 250 F.3d at 196–97 (holding that a public employee can state a claim for violation of free speech rights while performing work duties).
he conducted a routine investigation into an automotive scam implicating other investigators. In rejecting the argument of the prosecutor’s office that statements made in the course of carrying out employment duties received no constitutional protection, the court noted that the speech condemning the other investigators fit squarely into Connick’s public concern test, which emphasized the importance of bringing to light breaches of the public trust.

Similarly, in 2003, in Rodgers v. Banks, the Sixth Circuit Court of Appeals held that an employee could engage in protected speech as a citizen while performing employment duties. The court expressly rejected the district court’s finding that a letter written by the plaintiff could not pass the Connick public concern test because it was written in her capacity as an employee and not in her role as a citizen. Instead, the court explained that the Connick analysis requires a determination whether an employee was “speaking as a citizen (albeit in the employee role) [rather than] speaking as an employee for personal interest.” Rodgers decided that a proper focus under Connick is on the purpose of the speech as opposed to the role of the speaker in saying it. Again, a circuit court rejected the adoption of a per se rule.

Several other circuits have impliedly rejected a per se rule. For example, in 1999, in Lewis v. Cowen, the Second Circuit Court of Appeals held that a state lottery worker’s refusal to propose certain changes to the game was a matter of public concern, but that the state’s interest in efficiency outweighed the employee’s personal interest in the speech. By allowing the lottery worker’s speech made within his official capacity to pass the threshold Connick public concern test, the Second Circuit impliedly accepted the idea that public employees have protected speech rights within the scope of their employment duties.

Similarly, in 1998, in Dill v. City of Edmond, the Tenth Circuit Court of Appeals held that a police officer’s statement that he believed certain determinations about the facts of a homicide investiga-

---

124 Id. at 200.
125 Id. at 196–97; see also Connick, 461 U.S. at 147.
126 344 F.3d at 598–99.
127 Id. at 598.
128 Id. at 599.
129 Id.
130 See id.
131 See infra notes 132–36 and accompanying text.
132 165 F.3d at 164–65.
133 See id.
tion to be untrue was a matter of public concern even though the investigation was a part of his employment duties. Likewise, in 1996, in *Fikes v. City of Daphne*, the Eleventh Circuit Court of Appeals also held that a police officer’s statements regarding improprieties within the department touched on matters of public concern, regardless of the fact that the statements were made in the course of the officer’s duties. By acknowledging that these employees had constitutionally protected rights in carrying out employment tasks, these circuit courts impliedly rejected the per se rule.

C. Federal Circuit Courts of Appeals Using Other Approaches to a Per Se Rule

Unlike federal appeals courts that have expressly or impliedly rejected the per se rule, the Fifth Circuit Court of Appeals acknowledges a public-private distinction in determining whether public employee speech reaches matters of public concern under *Connick*. Yet in 2000 in *Kennedy v. Tangipahoa Parish Library Board of Control*, the Fifth Circuit held that even when a speaker has mixed motives in his or her role as a speaker, a statement still can pass the *Connick* public concern test. Therefore, in *Kennedy*, where a library worker suggested a new security policy in response to a rape on the premises, the court held the speech was still protected because the employee was speaking as both an employee and a citizen.

The Seventh Circuit Court of Appeals also has adopted a unique approach to the proposed per se rule. In 2001, in *Gonzalez v. City of Chicago*, the court held that employees’ routine reports were not of public concern under *Connick* because they formed part of a constant duty to fill out paperwork. In 2002, in *Delgado v. Jones*, however, the court narrowed the holding of *Gonzalez* to address only the most routine work tasks, where no evidence of public motivation exists. In *Delgado*, this meant a police officer’s reports detailing impropriety

---

134 155 F.3d at 1202–03.
135 79 F.3d at 1084.
136 See supra notes 132–35 and accompanying text.
137 See *Kennedy v. Tangipahoa Parish Library Bd. of Control*, 224 F.3d 359, 366 (5th Cir. 2000) (holding that a free speech right can exist where a public employee speaks both as a citizen and as an employee on matters of public concern while performing job tasks).
138 *Id.*
139 *Id.* at 376.
141 *Gonzalez v. City of Chi.*, 239 F.3d 939, 941 (7th Cir. 2001).
142 *Delgado*, 282 F.3d at 519.
within the department were protected.\textsuperscript{143} The court explained that \textquotedblleft Gonzalez is limited to routine discharge of assigned functions, where there is no suggestion of public motivation.	extquotedblright\textsuperscript{144} Although the \textit{Delgado} court expressly stated that it rejected a per se rule against speech rights in the course of employment duties, the \textit{Gonzalez} precedent still creates a per se rule preventing First Amendment protection for speech in the most routine job functions.\textsuperscript{145} Overall, then, the majority of circuits have expressly or impliedly rejected a per se rule.\textsuperscript{146}

\textbf{D. Policy Reasons for Rejecting the Per Se Rule}

The \textit{Ceballos} court and others opposed to a per se rule raise two policy objections.\textsuperscript{147} First, a per se rule would provide diminished protection for the whistleblower.\textsuperscript{148} Second, a per se rule would chill valuable speech.\textsuperscript{149}

The first policy objection to a per se rule is that such a rule would adversely affect the status of the whistleblower.\textsuperscript{150} Discouraging public employees from unveiling scandal, corruption, or illegal acts within the government defies reason, according to the \textit{Ceballos} court, because such matters are always of great public concern.\textsuperscript{151} In particular, the court reasoned that under the per se rule formulation, which distinguishes between speech by employees acting as citizens and speech by employees acting as employees, protecting only speech by citizens would encourage public employees to speak to the media or other indirect sources to seek redress—an indirect approach to resolution—instead of reporting the improper doings to superiors.\textsuperscript{152} The \textit{Ceballos} court also noted that the \textit{Connick} ruling draws no distinction between internal and external speech when determining whether public em-

\begin{itemize}
\item \textsuperscript{143} \textit{Id.}; see also \textit{Gonzalez}, 239 F.3d at 941. Only the most rote tasks that were of little public concern did not receive First Amendment protection after \textit{Delgado}. 282 F.3d at 519.
\item \textsuperscript{144} \textit{Delgado}, 282 F.3d at 519 (distinguishing \textit{Gonzalez}).
\item \textsuperscript{145} \textit{Id.}
\item \textsuperscript{146} See \textit{Rodgers}, 344 F.3d at 598–99; \textit{Baldassare}, 250 F.3d at 196–97; \textit{Lewis}, 165 F.3d at 164; \textit{Dill}, 155 F.3d at 1202–03; \textit{Fikes}, 79 F.3d at 1084. The First and Eighth Circuit Courts of Appeals have not addressed the issue sufficiently to discern their stance on the legality of a per se rule. \textit{See generally} Schilcher v. Univ. of Ark., 387 F.3d 959 (8th Cir. 2004); Rivera-Jiminez v. Pierluisi, 362 F.3d 87 (1st Cir. 2004).
\item \textsuperscript{147} See \textit{Ceballos}, 361 F.3d at 1176; \textit{Urofsky}, 216 F.3d at 439 (Murnaghan, J., dissenting).
\item \textsuperscript{148} See \textit{infra} notes 150–56 and accompanying text.
\item \textsuperscript{149} See \textit{infra} notes 157–59 and accompanying text.
\item \textsuperscript{150} \textit{Ceballos}, 361 F.3d at 1176.
\item \textsuperscript{151} \textit{See id.}
\item \textsuperscript{152} \textit{Id.}
ployee speech touches on matters of public concern. Therefore, articulating a test that encourages external whistleblowing and discourages internal whistleblowing adds a distinction to public employee speech unacknowledged by the Supreme Court. Furthermore, the court reasoned, whistleblower statutes provide only some protection for government employees who reveal wrongdoing. Although such statutes often allow for whistleblowing against the government, the First Amendment can more affirmatively protect what its very language commands: freedom from the suppression of the ability to seek redress from that government.

Finally, those opposing the per se rule reason that adopting such a rule will produce a general chilling effect on speech. This rationale reflects the policy consideration that because public employees generally have insight into the functioning of government, their speech is inherently valuable. Together with concerns about protection for whistleblowers, then, the potential chilling effect on speech discourages the adoption of a per se rule.

IV. ANALYSIS: THE PER SE RULE WOULD BE CONTRARY TO THE PURPOSE OF THE FIRST AMENDMENT

In addition to the previously discussed criticisms of the proposed per se rule precluding free speech protection for public employees performing regular job duties, courts must reject the rule because it is contrary to the purposes of the First Amendment. First, a per se rule could pose practical difficulties for professors in public universities and limit academic inquiry. Furthermore, a per se rule could result in an increase of impermissible viewpoint discrimination.

153 See id.
154 See id.
156 See U.S. Const. amend. I; Ceballos, 361 F.3d at 1174.
157 See Urofsky, 216 F.3d at 439 (Murnaghan, J., dissenting).
158 See id.
159 See supra notes 150–58 and accompanying text.
160 See infra notes 164–212 and accompanying text.
161 See Urofsky v. Gilmore, 216 F.3d 401, 439 (4th Cir. 2000) (Murnaghan, J., dissenting); see also infra notes 164–84 and accompanying text.
162 See Rosenthal, supra note 16, at 542; see also infra notes 185–212 and accompanying text. Content discrimination is the practice of forbidding speech because it addresses a certain topic. Rosenthal, supra note 16, at 540. Content discrimination can lead to viewpoint discrimination, which is the practice of forbidding speech that takes a standpoint disfavored by the suppressor of the speech. Id. at 542.
Additionally, the justifications for the per se rule are contrary to existing First Amendment caselaw.\textsuperscript{163}

\textbf{A. The Issue of Speech in Public Universities}

A per se rule is contrary to the purposes of the First Amendment because it would negatively affect academic inquiry in public universities.\textsuperscript{164} Professors at public educational institutions are employees of the state and as such would be subject to a per se rule adopted by courts.\textsuperscript{165} Allowing the per se rule in this context is problematic for two reasons.\textsuperscript{166} Professors often speak both as citizens and as employees, rendering it difficult as a practical matter to determine the role in which the professor speaks in a particular instance.\textsuperscript{167} Additionally, allowing such a rule could impede academic freedom.\textsuperscript{168}

Adopting the per se rule would create practical difficulties for courts, which would be forced to determine whether a professor spoke as an employee or as a citizen in an academic context.\textsuperscript{169} The per se rule adopted by the Fourth Circuit in \textit{Urofsky v. Gilmore} requires such a determination before a court can inquire into whether the subject of the speech involved matters of public concern.\textsuperscript{170} So, under the per se rule, courts must determine in which role, citizen or employee, a professor speaks in deciding if his or her speech is constitutionally protected.\textsuperscript{171} This determination is practically difficult because professors frequently publish articles and books, some in the name of their respective public employers and some not.\textsuperscript{172} They often make presentations at different events, participate in speaking engagements, and have scholarly discussions with colleagues.\textsuperscript{173} Often it is difficult to classify the role in which a professor speaks and whether the topic is of personal or employment interest.\textsuperscript{174} Moreover, if courts determine that all these activities occur within the professor’s role as an employee, all such activities would lose free speech protection un-

\textsuperscript{163} See infra notes 213–45 and accompanying text.
\textsuperscript{164} See \textit{Urofsky}, 216 F.3d at 439 (Murnaghan, J., dissenting).
\textsuperscript{165} See id. at 438–39.
\textsuperscript{166} See infra notes 169–84 and accompanying text.
\textsuperscript{167} See infra notes 169–76 and accompanying text.
\textsuperscript{168} See infra notes 177–84 and accompanying text.
\textsuperscript{170} See 216 F.3d at 420.
\textsuperscript{171} See id.
\textsuperscript{172} See id. at 439 (Murnaghan, J., dissenting).
\textsuperscript{173} See id..
\textsuperscript{174} See id. at 420 (majority opinion).
under the per se rule.\textsuperscript{175} Thus, it is practically difficult to apply the per se rule in the academic environment.\textsuperscript{176}

Besides such practical difficulties, a per se rule affecting public school professors could impede academic freedom.\textsuperscript{177} The Supreme Court recognizes that academic freedom is a special concern of the First Amendment.\textsuperscript{178} Professors and others involved in academia are often a large source of intellectual discourse because they frequently comment upon and criticize various governmental operations.\textsuperscript{179} If a per se rule is adopted, a professor’s scholarly discourse could lead to retaliatory action, with no First Amendment recourse, should a professor’s superior disagree with the professor’s scholarship.\textsuperscript{180} For example, under the per se rule, an economics professor would lack First Amendment protection in publishing an article criticizing current U.S. fiscal policy.\textsuperscript{181} Such a potential result counsels against the adoption of the per se rule because public university professors have unique knowledge of their areas of specialization and can contribute significantly to the debate in their fields.\textsuperscript{182} If courts adopt the per se rule, open debate on scholarly topics could diminish for fear of reprisal.\textsuperscript{183} The per se rule would thus eradicate free speech protections in one of the places where it is most important to protect them—public universities.\textsuperscript{184}

**B. The Risk of Content and Viewpoint Discrimination**

In addition to the risk of impeding academic freedom, the per se rule should be rejected because it increases the risk of impermissible viewpoint discrimination, which is the practice of forbidding speech because it expresses a standpoint disfavored by the suppressor of the speech.\textsuperscript{185} Whenever speech is restricted, the restriction constitutes some form of content discrimination because it disallows a certain

\begin{itemize}
\item \textsuperscript{175} See Urofsky, 216 F.3d at 420.
\item \textsuperscript{176} See supra notes 169–75 and accompanying text.
\item \textsuperscript{177} See infra notes 178–84 and accompanying text.
\item \textsuperscript{178} See supra notes 178–84 and accompanying text.
\item \textsuperscript{179} See Keyishian v. Bd. of Regents, 385 U.S. 589, 603 (1967) (holding unconstitutional a requirement that university faculty certify that they were not members of the Communist Party and recognizing a commitment to protect academic freedom).
\item \textsuperscript{180} See Urofsky, 216 F.3d at 439 (Murnaghan, J., dissenting).
\item \textsuperscript{181} See id.
\item \textsuperscript{182} See id.
\item \textsuperscript{183} See id.
\item \textsuperscript{184} See Urofsky, 216 F.3d at 439 (Murnaghan, J., dissenting).
\item \textsuperscript{185} See Rosenthal, supra note 16, at 542.
\end{itemize}
category of speech. For example, forbidding employees from engaging in sexual harassment in the workplace might be an acceptable form of content discrimination. Given that the Supreme Court in *Pickering v. Board of Education* held that a public employee receives free speech protection only if the employee’s interest in the speech outweighs the employer’s interest in promoting efficiency, content discrimination is permissible under *Pickering* if used to promote workplace efficiency.

Although some forms of content discrimination can be permissible, dangers of viewpoint discrimination can arise. For example, a government employer might permissibly forbid political conversations during departmental meetings for the sake of efficiency—a restriction that constitutes content discrimination. If the employer forbids only conversations about the Democratic Party, however, such a restriction constitutes viewpoint discrimination. Viewpoint discrimination is dangerous because it can be used to perpetuate the domination of majority views and suppress valuable speech.

The adoption of the per se rule would increase the risk of impermissible viewpoint discrimination in governmental workplaces. A proponent of the per se rule might argue that by precluding speech rights for employees performing work tasks, only permissible content discrimination exists concerning topics unrelated to assigned work. This justification is grounded in the *Pickering* concern of ensuring that government employers have latitude to manage and promote efficiency. Such an argument is misguided, however, because most likely the adoption of a per se rule would deter only some public employees from criticizing workplace operations or engaging in other speech, and many public employees would still speak freely despite their lack of First Amendment protection.

---

186 See id.
187 See id.
188 See 391 U.S. 563, 568 (1968).
190 See id.
191 See id.
192 See id.
193 See infra notes 194–212 and accompanying text.
194 See *Pickering*, 391 U.S. at 568. For example, a public employer might hypothetically argue that it is mere content discrimination to punish employees who mention politics within work assignments, grounding its argument for permissibility in the employer’s right under *Pickering* to promote workplace efficiency. See id.
195 See id.
196 See *Urofsky*, 216 F.3d at 439 (Murnaghan, J., dissenting).
would allow government employers simply to suppress the viewpoints they disfavor by arbitrarily reprimanding employees for speech with which the employers disagree.\textsuperscript{197} Government employers could engage in this type of viewpoint discrimination without fear of repercussion under the per se rule because such a rule leaves public employees with no recourse in courts.\textsuperscript{198} \textit{Pickering} authorizes government employers to regulate workplace efficiency, but the Supreme Court does not provide a license to do so by punishing only disagreeable ideas.\textsuperscript{199}

An increased risk of viewpoint discrimination within governmental workplaces is problematic because governmental interference within the political and moral realms could result.\textsuperscript{200} If government employers can suppress speech that individual managers deem unfavorable, issues regarding the type of speech suppressed could arise.\textsuperscript{201} Of course, concern arises that whistleblowing or other speech alleging mismanagement would be suppressed.\textsuperscript{202} Additionally, the danger exists that government employers would encourage certain political affiliations.\textsuperscript{203} Such a danger is real; for example, in \textit{Connick v. Myers}, the plaintiff received a reprimand for distributing a questionnaire asking other employees if pressure to conform to certain political views existed within the government workplace.\textsuperscript{204} Moreover, the danger exists that certain majority views on moral or controversial issues would be encouraged through the suppression of minority opinions.\textsuperscript{205} For example, government employers could allow speech in a law enforcement office recommending or supporting the death penalty and could reprimand workers who take a contrary view.\textsuperscript{206}

Exceptional dangers to free speech arise where government employers are able to suppress minority views through reprimand but can allow speech on viewpoints with which they agree.\textsuperscript{207} In its virtual endorsement of the per se rule, the \textit{Urofsky} court reasoned that speech rights should not protect public employees carrying out employment tasks because onlookers could perceive the speech as an

\textsuperscript{197} See id.

\textsuperscript{198} See id.

\textsuperscript{199} See 391 U.S. at 568.

\textsuperscript{200} See id.

\textsuperscript{201} See id.

\textsuperscript{202} See infra notes 226–36 and accompanying text.

\textsuperscript{203} See \textit{Connick}, 461 U.S. 138 at 142.

\textsuperscript{204} See id.

\textsuperscript{205} See Rosenthal, \textit{supra} note 16, at 542.

\textsuperscript{206} See id.

\textsuperscript{207} See id.
official state view. In practice, however, if employers can allow only speech they deem favorable without fear of First Amendment claims due to the per se rule, creating an official state view on certain issues becomes easier. One viewpoint could easily predominate in a government office if the decisionmakers in that office punish those who engage in deviating speech—a result that essentially itself creates an “official” view. An official or quasi-official state view on any political or moral issue is contrary to the purpose of the First Amendment. For all of these reasons, the per se rule could result in impermissible viewpoint discrimination.

C. The Lack of Justification for a Per Se Rule

In addition to impeding academic freedom and allowing impermissible viewpoint discrimination, policy considerations cited by the Urofsky court and those endorsing the per se rule logically fail as well. They contend that courts should adopt the rule for three major reasons. First, public employees speak for the state. Second, speech rights would be inconsistent with whistleblower statutes. Finally, the rule would constitutionalize performance of everyday work duties. Each of these arguments misconstrues the Connick public concern test and the Pickering balancing test.

The Urofsky view rests on the idea that public employees should not have free speech protection in carrying out employment duties because in fulfilling these duties, employees are speaking for the state. This position, however, ignores the factors that the U.S. Supreme Court articulated for performing the Connick public concern test. Under Connick, in determining whether employee speech touches on matters of public concern, courts must analyze the “con-

---

208 See 216 F.3d at 408 n.6.
209 See id.
210 See id. at 407–08.
212 See supra notes 185–211 and accompanying text.
213 See infra notes 214–45 and accompanying text.
214 See supra notes 93–105 and accompanying text.
215 See infra notes 219–25 and accompanying text.
216 See infra notes 226–36 and accompanying text.
217 See infra notes 237–45 and accompanying text.
218 See infra notes 219–45 and accompanying text.
219 See Ceballos v. Garcetti, 361 F.3d 1168, 1189 (9th Cir. 2004) (O’Scannlain, J., specially concurring), cert. granted, 125 S. Ct. 1395 (2005); Urofsky, 216 F.3d at 408 n.6.
220 See 461 U.S. at 147–48.
tent, form, and context” of the allegedly protected speech.\textsuperscript{221} The purpose of these factors is to provide courts with criteria for determining as a matter of law what speech addressed matters of public concern.\textsuperscript{222} The premise of a multi-factored test is that in the public employment context, the entire set of circumstances dictates whether or not speech is worthy of constitutional protection.\textsuperscript{223} For example, although an Internal Revenue Service employee speaking to the public about new tax policies might be a representative of the state while carrying out that employment task, a public university professor drafting a scholarly article criticizing new tax policies probably represents his or her own views and does not represent the state even though the professor is performing regular employment duties.\textsuperscript{224} Because the \textit{Connick} Court recognized that factual circumstances differ in public employee free speech cases, it also implicitly acknowledged that an employee can speak as a citizen and as an employee at the same time, not only as one or the other.\textsuperscript{225}

In addition to misconstruing \textit{Connick} by reasoning that public employees necessarily speak on behalf of the state while performing job tasks, supporters of the \textit{Urofsky} view also fail in arguing that affording free speech rights to public employees performing job tasks is inconsistent with whistleblower statutes.\textsuperscript{226} This position fails on two counts because whistleblower statutes provide mainly procedural protections, and some valuable speech does not constitute whistleblowing and thus is not covered by whistleblower statutes.\textsuperscript{227}

Whistleblower statutes do not exist because public employees lack free speech protection to report wrongdoing, as supporters of the per se rule might suggest.\textsuperscript{228} Instead, in \textit{Waters v. Churchill}, the Supreme Court clarified that absent a strong showing of workplace disruption, most employee speech receives First Amendment protection, especially because public employees are often in a unique position to comment upon governmental operations because of their insider

\textsuperscript{221} \textit{Id.}
\textsuperscript{222} See \textit{id.}
\textsuperscript{223} See \textit{id.}
\textsuperscript{224} See \textit{id.}
\textsuperscript{225} See 461 U.S. at 147–48.
\textsuperscript{226} See Ceballos, 361 F.3d at 1192 (O’Scannlain, J., specially concurring). Proponents of the per se rule reason that if the First Amendment protected employees in these circumstances, no need for whistleblower statutes would exist. See \textit{id.}
\textsuperscript{227} See \textit{infra} notes 228–36 and accompanying text.
\textsuperscript{228} See \textit{Waters v. Churchill}, 511 U.S. 661, 674 (1996).
viewpoint. Legislatures may choose to provide supplemental protections in whistleblower statutes out of respect for the values underlying the First Amendment and to prevent adverse employment actions taken against public employees who exercise speech rights. Whistleblower statutes do not create a new right to speak about governmental wrongdoing; rather, they provide employees who exercised their right with additional guarantees against the consequences of doing so. The result is procedural protection, not the creation of a new free speech right that did not previously exist.

Additionally, the per se rule could thwart constitutional protection for speech that does not constitute whistleblowing but still engenders valuable discourse, although it is part of the public employee’s job duties. For example, although a police officer reporting that another officer used excessive force on an arrestee might constitute whistleblowing, a police officer writing a report giving the number of instances of officers shooting arrestees and suggesting the number is high might not constitute whistleblowing because it is not necessarily unveiling wrongdoing. Nonetheless, such speech, which possesses the potential to improve governmental operations, should receive First Amendment protection even if the speech does not blow the whistle on wrongdoing. Overall, the mere existence of whistleblowing statutes does not negate the need for First Amendment protection for public employees carrying out job duties.

In addition to failing in the position that whistleblower statutes demonstrate the needlessness of First Amendment protection for public employees performing job tasks, the Urofsky court and those supporting its virtual adoption of the per se rule also fail in concluding that providing constitutional protection will constitutionalize everyday work duties. This argument attempts to substitute an absolute abolition on fundamental free speech rights where the Pickering test already allows leniency for government employers to control speech while still providing First Amendment protection. Supporters of the

229 See id.
230 See id.
232 See id.
234 See id.
235 See Waters, 511 U.S. at 674.
236 See supra notes 226–35 and accompanying text.
237 See 216 F.3d at 408.
238 See 391 U.S. at 568.
per se rule assume that by allowing constitutional protection for some employee speech contained in work tasks, free speech issues will pervade government workplaces.\textsuperscript{239} The Supreme Court clearly acknowledges, however, that government employers must enjoy great deference in managing operations, such as in forbidding the use of offensive language.\textsuperscript{240} For this very reason, the \textit{Pickering} test balances the government interest in workplace efficiency against the speech interests of public employees.\textsuperscript{241}

Thus, it is nonsensical to distinguish, as the \textit{Urofsky} court does by adopting the per se rule, between public employees speaking as citizens and those speaking as employees, in an attempt to preserve employer power to regulate speech because such power already exists in the \textit{Pickering} balancing test.\textsuperscript{242} Allowing free speech protection for public employees performing regular job duties will not constitutionalize all work tasks.\textsuperscript{243} Instead, the \textit{Urofsky} viewpoint fails in each of the three rationales offered to justify the adoption of the per se rule.\textsuperscript{244} These arguments misconstrue the proper application of the \textit{Pickering} balancing test and the \textit{Connick} public concern test; therefore, courts should reject the per se rule when analyzing public employee speech under the First Amendment.\textsuperscript{245}

\textbf{Conclusion}

Adopting a per se rule precluding free speech rights for public employees performing regular job duties would be contrary to the purposes of the First Amendment. The rule would infringe upon academic freedom because professors at public universities engage in academic discourse as a regular employment responsibility. Moreover, such a rule increases the risk of viewpoint discrimination because it allows employers to punish only speech they find personally unfavorable and leaves employees with no cause of action when reprimanded for their speech. Additionally, no logical justification for the per se rule exists. Public employees do not always speak on behalf of the state when carrying out job tasks, and whether speech is protected is determined through the circumstance-based inquiry in \textit{Connick}. Whis-

\textsuperscript{239} See \textit{Urofsky}, 216 F.3d at 408.
\textsuperscript{240} \textit{Waters}, 511 U.S. at 672.
\textsuperscript{241} See 391 U.S. at 568.
\textsuperscript{242} See \textit{id}.
\textsuperscript{243} See \textit{id}.
\textsuperscript{244} See supra notes 213–43 and accompanying text.
\textsuperscript{245} See supra notes 213–43 and accompanying text.
tleblower statutes merely supplement existing rights; they do not render these rights ineffectual. Moreover, employer interests in suppressing disruptive speech are weighed in the *Pickering* balancing test, quelling concerns that speech rights constitutionalize everyday job tasks. Because of the lack of justification for a per se rule precluding free speech rights for public employees performing regular job duties, courts must reject the adoption of the per se rule.

Marni M. Zack